

**Der Zusammenhang zwischen akkurater Selbsteinschätzung und Performanz
im wirtschaftswissenschaftlichen Kontext**

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften der
Universität Mannheim

Vorgelegt von
Dipl.-Hdl. Volker Rohr
Adliswil

Dekan: Professor Dr. Dieter Truxius

Referent: Professor Dr. Hermann G. Ebner

Korreferent: Professor Dr. Jürgen Seifried

Tag der mündlichen Prüfung: 21.02.2018

Inhaltsverzeichnis

<i>Inhaltsverzeichnis</i>	<i>I</i>
<i>Abbildungsverzeichnis</i>	<i>IV</i>
<i>Tabellenverzeichnis.....</i>	<i>VI</i>
<i>Abkürzungsverzeichnis.....</i>	<i>VIII</i>
Kapitel 1 Einleitung.....	1
1.1 Die Bedeutung der Selbsteinschätzungen im studentischen Kontext.....	1
1.2 Zielsetzung und Aufbau der Arbeit.....	5
Kapitel 2 Konzeptuelle Grundlagen: Selbsteinschätzungen in der Metakognitionsforschung	6
2.1 Die Bedeutung der Selbsteinschätzungen in der Metakognitionsforschung	6
2.2 Die Genauigkeit der Selbsteinschätzungen von Lernenden.....	10
2.3 Effekte der Initiierung von Selbsteinschätzungen	14
Kapitel 3 Empirische Befunde zum Zusammenhang zwischen Akkuratheit der Selbsteinschätzung, selbstreguliertem Lernen und der Performanz	16
3.1 Studienrecherche.....	16
3.2 Studienergebnisse	20
3.2.1 Effekte (in-)akkurater Selbsteinschätzungen auf Selbstregulation und Performanz.....	20
3.2.2 Effekte (in-)akkurater Selbsteinschätzungen auf die Performanz	29
3.2.3 Effekte (in-)akkurater Selbsteinschätzungen auf weitere lernbedeutsame Konstrukte.....	37
3.3 Diskussion der Forschungsbefunde	40
3.3.1 Zusammenfassung der Ergebnisse.....	40
3.3.2 Implikationen für die weitere Forschung.....	44
Kapitel 4 Fragestellung und Hypothesen	48

Kapitel 5	Untersuchungsdesign	51
5.1	Anlage der Untersuchung	51
5.2	Designkomponenten des ersten Messzeitpunktes	53
5.2.1	Stichprobe	53
5.2.2	Instrument	55
5.2.2.1	Darstellung und Klassifizierung der Testaufgaben	55
5.2.2.2	Entwicklung von Items zur Erfassung der Selbsteinschätzung	65
5.2.2.3	Entwicklung von Items zur Erfassung der eingesetzten Lernstrategien	67
5.2.2.4	Entwicklung von Items zur Erfassung des Zeitaufwands	73
5.2.2.5	Allgemeine Angaben	73
5.2.3	Ablauf der Datenerhebung	74
5.3	Designkomponenten des zweiten Messzeitpunktes	75
5.3.1	Stichprobe	75
5.3.2	Instrument	78
5.3.3	Ablauf der Datenerhebung	78
5.4	Datenauswertung	79
Kapitel 6	Ergebnisse	81
6.1	Ausprägung der Selbsteinschätzungen und Performanz	81
6.1.1	Selbsteinschätzungen und Performanz der Gesamtstichprobe	81
6.1.2	Selbsteinschätzungen und Performanz der Teilstichprobe	88
6.2	Befunde zur Akkuratheit der Selbsteinschätzung	93
6.2.1	Ausprägung akkurater Selbsteinschätzungen an Messzeitpunkt I	93
6.2.2	Ausprägung akkurater Selbsteinschätzungen an Messzeitpunkt II	99
6.2.3	Exkurs: Einflussfaktoren auf die Akkuratheit der Selbsteinschätzungen ..	103
6.3	Zusammenhänge zwischen akkurater Selbsteinschätzung und Perfor- manz	106
6.4	Lernstrategien und Lernzeit in Abhängigkeit von der Akkuratheit der Selbsteinschätzung	118
Kapitel 7	Diskussion und Ausblick	125
7.1	Zusammenfassung zentraler Ergebnisse	125
7.2	Diskussion der zentralen Ergebnisse	130
7.3	Limitationen der Untersuchung und Implikationen für die Forschung	136

<i>Literaturverzeichnis</i>	141
<i>Anhang</i>	151

Abbildungsverzeichnis

Abbildung 1	Phasen der Selbstregulation nach Zimmerman & Campillo (2003, S. 239).....	3
Abbildung 2	Lernertypologien nach dem Ansatz des Confidence-Based Assessment and Learning	4
Abbildung 3	Metakognition nach Flavell (1979).....	6
Abbildung 4	Kontroll- und Monitoringprozesse nach Nelson & Narrens (1990)	8
Abbildung 5	Einflussfaktoren auf die Akkuratheit der Selbsteinschätzung	13
Abbildung 6	Initiierungszeitpunkte der Selbsteinschätzung.....	17
Abbildung 7	Potentielle Zusammenhänge zwischen Fehleinschätzung und Performanz.....	45
Abbildung 8	Anlage der Untersuchung	52
Abbildung 9	Revidierte Taxonomien kognitiver Prozesse nach Anderson & Krathwohl (2001).....	58
Abbildung 10	Kategoriensystem zur Analyse von Aufgaben (Maier et al., 2010).....	61
Abbildung 11	Klassifikationsmatrix in ULME nach Brand, Hofmeister, & Tramm (2005) sowie Hofmeister (2005).....	63
Abbildung 12	Vergleich der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt I und Messzeitpunkt II.....	82
Abbildung 13	Vergleich der Selbsteinschätzungen zwischen beiden Messzeitpunkten vor der Aufgabenbearbeitung	89
Abbildung 14	Vergleich der Selbsteinschätzungen zwischen beiden Messzeitpunkten nach der Aufgabenbearbeitung	90
Abbildung 15	Vergleich der Überzeugungen zur korrekten Bearbeitung der Aufgaben zwischen beiden Messzeitpunkten.....	91
Abbildung 16	Vergleich der an beiden Messzeitpunkten im Mittel erreichten Punkte ..	92
Abbildung 17	Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt I.....	95
Abbildung 18	Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt I	97
Abbildung 19	Akkuratheit der Überzeugungen an Messzeitpunkt I.....	98

Abbildung 20	Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt II.....	101
Abbildung 21	Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt II.....	102
Abbildung 22	Akkuratheit der Überzeugungen an Messzeitpunkt II	103
Abbildung 23	Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt I.....	107
Abbildung 24	Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt I	108
Abbildung 25	Akkuratheit der Überzeugungen an Messzeitpunkt I.....	109
Abbildung 26	Entwicklung der Performanz in Abhängigkeit von der Akkuratheit der Selbsteinschätzung an Messzeitpunkt I	115
Abbildung 27	Ausprägung der Lernstrategien zu Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Buchungsaufgaben) zu Messzeitpunkt I.....	121
Abbildung 28	Erwarteter Zusammenhang zwischen Fehleinschätzungen und Lernergebnis	135

Tabellenverzeichnis

Tabelle 1	Überblick über Studien der ersten Kategorie.....	20
Tabelle 2	Überblick über Studien der zweiten Kategorie.....	29
Tabelle 3	Überblick über Studien der dritten Kategorie.....	37
Tabelle 4	Übersicht der Fragestellungen und Hypothesen	50
Tabelle 5	Charakterisierung der Stichprobe zu Messzeitpunkt I.....	54
Tabelle 6	Übersicht der für die Befragung ausgewählten Aufgaben.....	57
Tabelle 7	Zuordnung der Items zur Selbsteinschätzung zu den äquivalenten Aufgaben.....	66
Tabelle 8	Übersicht der Subskalen zu den Lernstrategien nach Pintrich et al. (1991).....	69
Tabelle 9	Übersicht der Subskalen des LASSI (Weinstein, Zimmermann, & Palmer, 1988).....	70
Tabelle 10	Übersicht der Items zum Einsatz von Lernstrategien	72
Tabelle 11	Übersicht der Items zur Erfassung des Zeitaufwands.....	73
Tabelle 12	Charakterisierung der Stichprobe zu Messzeitpunkt II.....	76
Tabelle 13	Charakterisierung der Teilstichprobe.....	77
Tabelle 14	Auswertungsplan.....	80
Tabelle 15	Mittelwerte und Standardabweichungen der Selbsteinschätzungen	83
Tabelle 16	Übersicht über Performanzwerte an beiden Messzeitpunkten.....	87
Tabelle 17	Performanzwerte an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung an Messzeitpunkt I	113
Tabelle 18	Performanzwerte an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung an Messzeitpunkt I	114
Tabelle 19	Kreuztabelle für Fehleinschätzungen in Bezug auf MC-Aufgaben zu Messzeitpunkt I und Performanz bei MC-Aufgaben zu Messzeitpunkt II.....	116
Tabelle 20	Angabe realisierter Lernstrategien zu Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung zu Messzeitpunkt I	120

Tabelle 21	Zeitaufwand an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung zu Initiierungszeitpunkt I an Messzeitpunkt I...	124
Tabelle 22	Übersicht zur Akkuratheit der Selbsteinschätzung für die eingesetzten Aufgabentypen zu den verschiedenen Mess- und Initiierungszeitpunkten.....	127

Abkürzungsverzeichnis

<i>ANOVA</i>	<i>Analysis of Variance</i>
bspw.	beispielsweise
bzgl.	bezüglich
CJ	Confidence Judgement
<i>d</i>	<i>Effektstärke</i>
<i>df</i>	<i>degrees of freedom (Freiheitsgrade)</i>
d. h.	das heißt
ECTS	European Credit Transfer System
EoL	Ease of Learning-Judgement
ERIC	Education Resources Information Centre
et al.	et alii (und andere)
GoB	Grundsätze ordnungsmäßiger Buchführung
GuV	Gewinn und Verlust
HZB	Hochschulzugangsberechtigung
i. H. v.	in Höhe von
i. S.	im Sinne
i. V.	im Vergleich
IZP	Initiierungszeitpunkt
JoK	Judgement of Knowing
JoL	Judgement of Learning
KG	Kontrollgruppe
KMK	Kultusministerkonferenz
LASSI	Learning and Study Strategies Inventory
<i>M</i>	<i>Mean (arithmetisches Mittel)</i>
MAI	Metacognitive Awareness Inventory

MC	Multiple Choice
MMB	Multimedia Comprehension Battery
MSLQ	Motivated Strategies for Learning Questionnaire
MZ	Messzeitpunkt
N	<i>Stichprobenumfang</i>
n	<i>Teilstichprobenumfang</i>
n. s.	nicht signifikant
OHG	Offene Handelsgesellschaft
p	<i>Irrtumswahrscheinlichkeit</i>
r	<i>Korrelationskoeffizient</i>
SD	<i>Standard Deviation (Standardabweichung)</i>
t	<i>t-Wert (Prüfgröße)</i>
TG	Treatment-Gruppe
tlw.	teilweise
u. a.	unter anderem
ULME	Untersuchung von Leistungen, Motivation und Einstellungen
UStG	Umsatzsteuergesetz
vs.	versus
VWL	Volkswirtschaftslehre
w/f-Aussagen	wahr/falsch-Aussagen
wdh.	wiederholender
z. B.	zum Beispiel
β	<i>Regressionsparameter</i>
γ	<i>Korrelationsparameter</i>
χ^2	<i>Chi² (Prüfgröße)</i>

1 Einleitung

1.1 Die Bedeutung der Selbsteinschätzungen im studentischen Kontext

Der Studienabbruch ist ein Phänomen, das im deutschsprachigen Raum seit einiger Zeit verstärkt Gegenstand der Forschung ist (z. B. Blüthmann, Thiel, & Wolfgramm, 2011; Brandstätter, Grillich, & Farthofer, 2006; Wolter, Diem, & Messer, 2013). Aktuelle Studien zeigen, dass Studierende, die ihr Bachelor-Studium 2008/2009 an einer deutschen Universität begonnen haben, dieses zu 33 % ohne Abschluss beenden (Heublein, Richter, Schmelzer, & Sommer, 2014). Die Studienabbruchquoten variieren dabei in Abhängigkeit zur studierten Fächergruppe. Sie liegen für Fächer der Mathematik/Naturwissenschaften mit 39 % am höchsten, während sie in den Rechts-/Wirtschafts-/Sozialwissenschaften am niedrigsten (27 %) ausgeprägt sind (Heublein et al., 2014, S. 4). Bei der Eruierung der Ursachen für einen Studienabbruch wird oftmals die mangelnde Studierfähigkeit der Studierenden genannt. Konegen-Griener führte 2002 eine Befragung von 1.435 Professoren in Deutschland mit dem Ziel durch, die Studierfähigkeit der Studierenden aus ihrer Sicht zu erfahren. Hierzu wurden die Professoren um ihre Einschätzung hinsichtlich 45 kognitiver, persönlicher, sozialer und fachlicher Merkmale der Studierenden gebeten. Allgemein kommt Konegen-Griener (2002, S. 481) zum Ergebnis, dass nur jeder vierte Studierende aus Sicht der Professoren studierfähig ist. 40 % der Studierenden wird demnach eine mittlere Studierfähigkeit bescheinigt, jedem dritten Studierenden wird diese überhaupt nicht attestiert. Es fehle den Studierenden insbesondere an den kognitiven Fähigkeiten. Die persönlichkeitsbedingten Faktoren werden günstiger eingeschätzt, jedoch in der Mehrzahl auch eher kritisch gesehen. Köller (2014) fasst die Ergebnisse von Konegen-Griener (2002) derart zusammen, dass sowohl Intelligenz, Kreativität und sprachliche Kompetenzen als auch Motivation und Selbstregulation bedeutsame Dimensionen des Konstruktes Studierfähigkeit sind. Diese empirisch ermittelten Dimensionen sind nach Köller (2014) vergleichbar mit der von der Expertenkommission der Kultusministerkonferenz (KMK) formulierten Einschätzung, für die Studieneignung seien auch soziale Kompetenz, Selbststeuerung des Lernens und Eigenverantwortlichkeit maßgeblich (KMK, 1995).

Die kognitiven Fähigkeiten der Studierenden werden bei der Untersuchung der Studierfähigkeit häufig über die Abiturnote erfasst (z. B. Trapmann, Hell, & Weigand, 2007), die sich als ein Faktor mit recht hoher prädiktiver Kraft für die Vorhersage des Studienerfolgs erweist. Daneben rücken auch zunehmend die Fähigkeiten der Studierenden zur Selbstregulation in den Fokus der Forschungsbemühungen (z.B. Schmitz, 2001).

Baumert et al. (2000, S. 2) definieren Selbstregulation damit, „*in der Lage zu sein, Wissen, Fertigkeiten und Einstellungen zu entwickeln, die zukünftiges Lernen fördern und erleichtern und die – vom ursprünglichen Lernkontext abstrahiert – auf andere Lernsituationen übertragen werden können*“. In diesem Zusammenhang wurden Voraussetzungen formuliert, mit denen selbstreguliertes Lernen erreicht werden kann. Diese sehen neben der Vorbereitung des Lernens und der Durchführung der Lernhandlung auch eine Form der Bewertung der Lernleistung vor (Simons, 1992). Hierbei ist es wichtig, dass die Lernenden „sich selbst Rückmeldung über Lernprozess und –ergebnisse geben können“ sowie „Lernprozess und –ergebnisse realistisch bewerten können“ (Simons, 1992, S. 255).

Zimmerman (2000, 2002) konkretisiert die Definition von selbstreguliertem Lernen und kommt zum Schluss, dass Selbstregulation ein dynamisches Zusammenwirken von kognitiven, metakognitiven und motivationalen Aspekten des Lernens umfasst. Voraussetzung dabei ist, dass der Lernende aktiv in seinem Lernen involviert ist. Dies kommt im zyklischen Modell selbstregulierten Lernens von Zimmerman & Campillo (2003) zum Ausdruck, bei welchem drei Phasen der Selbstregulation dargestellt werden: *Forethought*, *Performance* und *Self-Reflection* (s. Abbildung 1). Im Rahmen der dem eigentlichen Lernen vorgelagerten *Forethought Phase* sind für den Lernenden die Aufgabenanalyse (Zielsetzung und Planung der einzusetzenden Lernstrategien) sowie die motivationalen Überzeugungen (abhängig von Zielorientierung, Interesse, Selbstwirksamkeitserwartung und erwarteter Performanz) von Bedeutung. In der anschließenden *Performance Phase* stehen die Selbstkontrolle und Selbstbeobachtung im Fokus. Lernstrategien werden eingesetzt und überwacht sowie die Konzentration und Motivation aufrechterhalten. In der abschließenden *Self-Reflection Phase* erfolgt die Bewertung des Lernergebnisses und des Lernprozesses durch den Lernenden, indem der Grad der Zielerreichung durch einen Vergleich der Performanz mit einem festgelegten Standard eingeschätzt wird. Diese Selbsteinschätzung bildet die Grundlage für den weiteren Lernprozess und das Setzen/Adaptieren von Lernzielen. Die Selbsteinschätzung des eigenen Wissensstandes ist demnach für den selbstregulierten Lernprozess von zentraler Bedeutung. Sie ist im Rahmen der *Self-Reflection Phase* das bindende Element zwischen den bereits zurückliegenden Lernprozessen und dem weiteren Fortgang des Lernens.

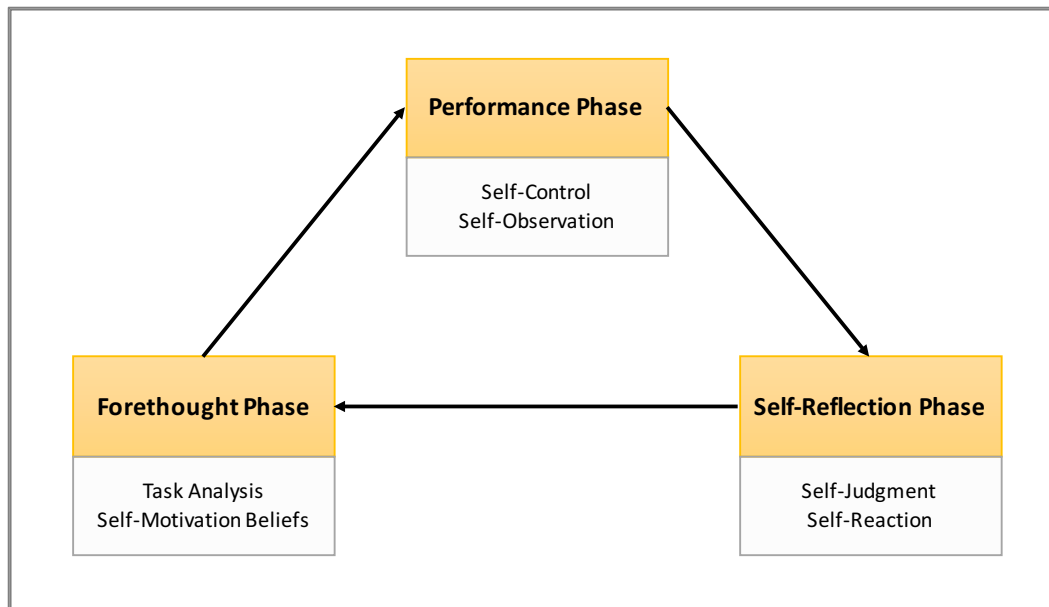


Abbildung 1: Phasen der Selbstregulation nach Zimmerman & Campillo (2003, S. 239)

In der Literatur wird für die Selbsteinschätzung eine Vielzahl von Begriffen verwendet, die häufig synonym betrachtet werden (Brown & Harris, 2013). Diese umfassen insbesondere *Self-Evaluation*, *Self-Assessment*, *Self-Reflection*, *Self-Grading*, *Self-Testing*, *Self-Score-Judgement* und *Self-Rating*. Gemein ist den Begrifflichkeiten, dass sie eine Einschätzung hinsichtlich der Güte des Gelernten durch den Lernenden selbst vorsehen. Klenowski (1995, S. 147) definiert *Self-Evaluation* als „*the evaluation or judgement of the ,worth‘ of one’s performance and the identification of one’s strengths and weaknesses with a view to improving one’s learning outcomes*“. Neben solchen Definitionsansätzen, die aus Erkenntnissen der Entwicklungspsychologie abgeleitet wurden, existieren auch solche, die ihren Ursprung in der pädagogisch-psychologischen Lernforschung haben. Beispielsweise geben Falchikov & Boud (1989) eine differenziertere Definition von *Self-Assessments*. Sie sehen hierin auch eine instruktionale Technik, mit der die Einbindung der Lernenden in den Prozess der Bewertung des eigenen Lernprozesses realisiert werden kann. Im Vordergrund steht dabei die Einschätzung der gezeigten Performanz. Mit dem Einsatz von *Self-Assessments* ist die Erwartung verbunden, eine aktivere Rolle der Lernenden in der Einschätzung des eigenen Lernprozesses zu bewirken.

Die Relevanz von Selbsteinschätzungen zeigt sich auch in Fragestellungen aus dem betriebswirtschaftlichen Kontext. In der betrieblichen Weiterbildung besteht eines der zentralen Probleme darin, die Behaltensleistung von in Trainings gelernten Inhalten so zu fördern, dass diese auch im Arbeitsleben angewandt werden können (Kim & Leaman, 2015). Eine Möglichkeit der Förderung besteht in der Berücksichtigung der Überzeugung der Angestellten, über relevantes

Wissen zur Problembewältigung im Arbeitsalltag zu verfügen. Das *Confidence-Based Assessment and Learning* setzt an der Fragestellung an, wie gemessen werden kann, ob Lernende oder Beschäftigte tatsächlich verstehen, was sie verstehen müssen, um erfolgreich zu lernen bzw. Arbeitsaufgaben bewältigen zu können (Adams & Ewen, 2009). Bei diesem Ansatz werden die Lernenden angehalten, während der Aufgabenbearbeitung die Überzeugung anzugeben, dass sie die Aufgabe korrekt lösen (Bruno, 1993). Auf Basis dessen lassen sich vier Typen von Lernenden unterscheiden (s. Abbildung 2).

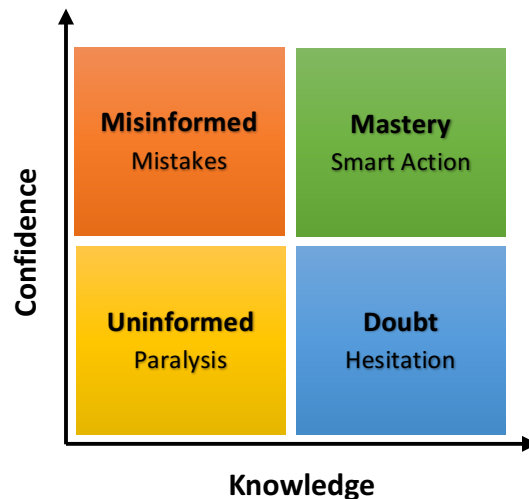


Abbildung 2: Lernertypologien nach dem Ansatz des *Confidence-Based Assessment and Learning*

Lernende mit hoher Überzeugung und geringem tatsächlichem Wissen entsprechen dem Typus des Überschätzers und neigen dem Ansatz des *Confidence-Based Assessment and Learning* folgend zu Fehlern. Dahingegen besteht bei Lernenden, die sich unterschätzen (hohes Wissen, geringe Überzeugung), die Gefahr des Zögerns bei der Erledigung von Aufgaben. Akkurate Selbsteinschätzungen können bei hohem Wissensstand zu einer guten Performanz, bei geringem Wissen jedoch zu einer „Lähmung“ im Lernprozess führen.

Diese Lernertypologien lassen sich auf den studentischen Kontext übertragen und können Aufschlüsse über Studienabbruchmotive liefern. Nach Heublein, Hutzsch, Schreiber, Sommer, & Besuch (2010) sind Leistungsprobleme und fehlende Studienmotivation ausschlaggebend für einen Studienabbruch. Neigen Studierende dazu, ihren eigenen Wissensstand zu überschätzen, könnte das ausbleibende Kompetenzerleben die Studienmotivation negativ beeinflussen. Es erscheint denkbar, dass eine möglichst akkurate Selbsteinschätzung an dieser Stelle förderlich für den Studienerfolg sein kann, da dem Wissensstand des Studierenden adäquate Lernaufgaben im selbstregulierten Lernprozess ausgewählt werden und die Lernzeit effektiv verteilt wird (Kicken, Brand-Gruwel, & van Merriënboer, 2008; Kostons, van Gog, & Paas, 2010, 2012).

1.2 Zielsetzung und Aufbau der Arbeit

Der in der Einleitung aufgezeigten Bedeutung von Selbsteinschätzungen folgend wird mit dieser Arbeit das Ziel verfolgt, in einem ersten Schritt die Selbsteinschätzung im universitären Kontext für Studierende in Studiengängen mit wirtschaftswissenschaftlicher Ausrichtung auf ihre Akkuratheit zu prüfen und darauf aufbauend in einem Folgeschritt mögliche Zusammenhänge zwischen akkurater Selbsteinschätzung und Selbstregulation sowie Performanz zu eruieren.

Zunächst erfolgt in Kapitel 2 eine vertiefende theoretische Einbettung der Thematik. Im Vordergrund steht dabei die Einordnung von Selbsteinschätzungen in das aus der Entwicklungspsychologie stammende Konzept der Metakognition. Hierbei kristallisieren sich über die Darstellung unterschiedlicher Formen der Selbsteinschätzung Kategorisierungsmöglichkeiten heraus, die für den weiteren Aufbau der Arbeit wegweisend sind. Flankierend hierzu werden zentrale Forschungsergebnisse zur Akkuratheit der Selbsteinschätzung sowie zu den Effekten der Initiierung von Selbsteinschätzungen im Lernprozess dargestellt. Es zeigt sich, dass die Beziehung zwischen akkurater Selbsteinschätzung und Performanz noch nicht hinreichend geklärt zu sein scheint. Aus diesem Grund wird in Kapitel 3 der Forschungsstand zum postulierten Zusammenhang eingehender dargestellt. Die Diskussion der Forschungsbefunde führt zu Implikationen für die weitere Forschung und ist handlungsleitend für das Vorgehen im Rahmen der empirischen Untersuchung. In Kapitel 4 erfolgt die Präzisierung der Fragestellung und die Darstellung der Hypothesen. Das Untersuchungsdesign wird in Kapitel 5 erläutert. Dabei liegt der Schwerpunkt in der Darstellung der Entwicklung des eingesetzten Instruments (s. Kapitel 5.2.2). Da das Instrument an beiden Messzeitpunkten gleichermaßen eingesetzt wurde, wird es im Rahmen der Beschreibung des Untersuchungsdesigns am zweiten Messzeitpunkt weniger ausführlich dargestellt. Das anschließende sechste Kapitel dient der Darstellung der Ergebnisse. Hierbei stehen nach einer deskriptiven Darstellung der Resultate (s. Kapitel 6.1) in Kapitel 6.2 die Ergebnisse zur Akkuratheit der Selbsteinschätzung im Vordergrund. Anschließend werden die Zusammenhänge zur Performanz (s. Kapitel 6.3) und der eingesetzten Lernstrategien (s. Kapitel 6.4) geprüft. In Kapitel 7 werden die Befunde diskutiert sowie auf Limitationen der Forschungsarbeit hingewiesen. Abschließend werden Implikation für Praxis und Forschung dargelegt.

2 Konzeptuelle Grundlagen: Selbsteinschätzungen in der Metakognitionsforschung

2.1 Die Bedeutung der Selbsteinschätzungen in der Metakognitionsforschung

In der Einleitung (s. Kapitel 1.1) wurde bereits die Bedeutung der Selbsteinschätzungen im selbstregulierten Lernprozess skizziert. Das Konzept des selbstregulierten Lernens hat seinen Ursprung in der pädagogisch-psychologischen Lernforschung (Hasselhorn & Labuhn, 2008). Weiterhin lässt sich die Thematik der Selbsteinschätzungen auch in das Konzept der Metakognition und somit aus einer entwicklungspsychologischen Perspektive theoretisch einordnen. Dabei betonen Hasselhorn & Labuhn (2008, S. 28), dass sich beide Konzepte weniger in den betrachteten „Phänomenen und Mechanismen menschlichen Verhaltens“ unterscheiden, sondern lediglich aus verschiedenen Forschungstraditionen entstanden sind.

Seit Beginn der 1970er-Jahre ist Metakognition ein wichtiger Bereich der entwicklungspsychologischen Forschung. Maßgeblich geprägt wurde die Forschung von den Arbeiten Flavells. Nach Flavell (1984, S. 23) wird unter Metakognition das „Wissen und Denken über kognitive Sachverhalte“ verstanden. Hasselhorn (1992) verweist darauf, dass sich in der Literatur die Unterscheidung zwischen einem deklarativen Wissensaspekt der Metakognition und einem exekutiven Kontrollprozess durchgesetzt hat. Dem folgend entstanden in der Vergangenheit verschiedene Klassifizierungsversuche hinsichtlich der Metakognition.

Flavell (1979) selbst unterscheidet zwischen metakognitivem Wissen (*metacognitive knowledge*), welches dem deklarativen Wissensaspekt zuzuordnen ist und metakognitiver Empfindung (*metacognitive experience*). Das metakognitive Wissen umfasst dabei Person-, Aufgaben- und Strategievariablen (s. Abbildung 3).

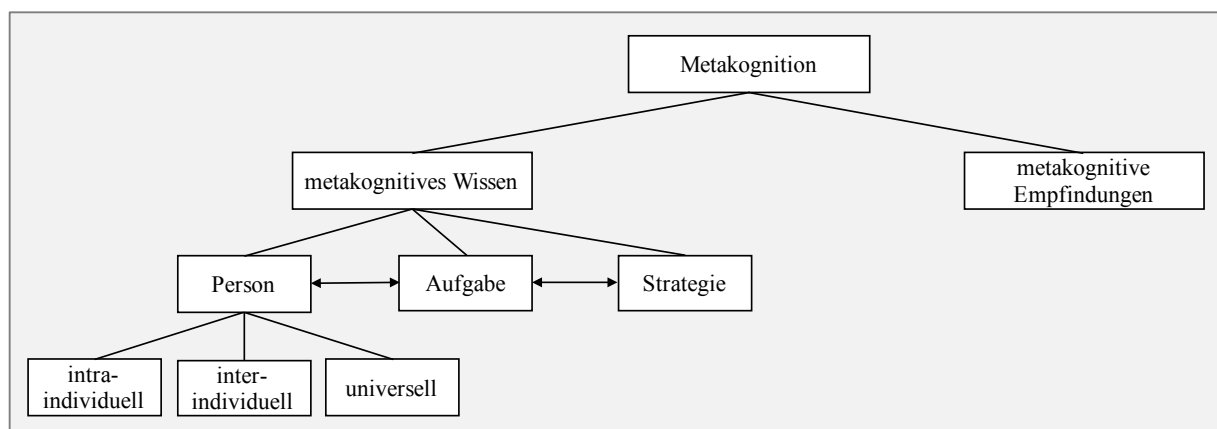


Abbildung 3: Metakognition nach Flavell (1979)

Die Personvariable wird nochmals nach intraindividuellem, interindividuellem und universellem Wissen ausdifferenziert. Intraindividuelles Wissen umfasst nach Flavell (1984) bspw. die Überzeugung einer Person, bestimmte Aufgaben gut und andere Aufgaben weniger gut bewältigen zu können. Es besteht also aus „Wissen und Annahmen über intraindividuelle Variationen hinsichtlich der eigenen [...] Fähigkeiten“ (Flavell, 1984, S. 24). Beim interindividuellen Wissen geht es um den Vergleich der eigenen Person mit anderen. So halten sich manche Personen bspw. für intelligenter als andere Personen. Das universelle Wissen umfasst „allgemeine Aspekte menschlichen Denkens“ (Flavell, 1984, S. 24).

Die Aufgabenvariable beinhaltet das Wissen über die Art der Aufgabe. Dies umfasst bspw. das Wissen über das kognitive Anspruchsniveau einer Aufgabe, welches die Auseinandersetzung mit der Aufgabe beeinflusst. Mit Strategievariablen umschreibt Flavell (1984, S. 25) das „Wissen über kognitive Strategien oder Prozeduren, um einen gegebenen Zustand zu verändern und um Ziele anzustreben“.

Metakognitive Empfindungen sind nach Flavell (1984) bewusste Empfindungen, die mit kognitiven Bemühungen in Verbindung stehen. Sie beziehen sich auf Tätigkeiten, die gerade ausgeführt werden. So können bspw. negative Empfindungen entstehen, wenn etwas nicht verstanden wird, was verstanden werden muss.

Hasselhorn (1992, S. 42) stellt ein integratives Klassifikationsschema der wichtigsten Subkategorien der Metakognition dar, in dem die von Flavell (1979) angeführten Variablen erweitert werden. Demnach umfasst Metakognition Systemisches Wissen, Epistemisches Wissen, Exekutive Prozesse (Kontrolle), Sensitivität für die Möglichkeiten kognitiver Aktivitäten und Metakognitive Erfahrungen bezüglich der eigenen kognitiven Aktivität.

Beiden Klassifizierungsansätzen gemein ist die Bedeutung der Selbsteinschätzung im Rahmen der Metakognition. Sie kommt bei Flavell (1979, 1984) insbesondere bei der Ausprägung des intra- und interindividuellen Wissens zum Tragen, während Selbsteinschätzungen im Klassifikationsschema von Hasselhorn (1992) besonders im Rahmen des epistemischen Wissens, welches u. a. das Wissen über Inhalte und Grenzen des eigenen Wissens umfasst, und der exekutiven Prozesse (Kontrolle) gewichtig sind. Hierunter wird die Planung, Überwachung und Steuerung des eigenen Lernprozesses subsumiert.

Eng verbunden mit dem Konzept der Metakognition und in Teilen auch synonym verwendet ist das von Flavell (1971) begründete Konzept des Metagedächtnisses (*Metamemory*), welches das

Wissen beschreibt, das eine Person über sein Gedächtnis besitzt. Nelson & Narens (1990) geben dem Konzept einen theoretischen Rahmen, der in Abbildung 4 dargestellt ist.

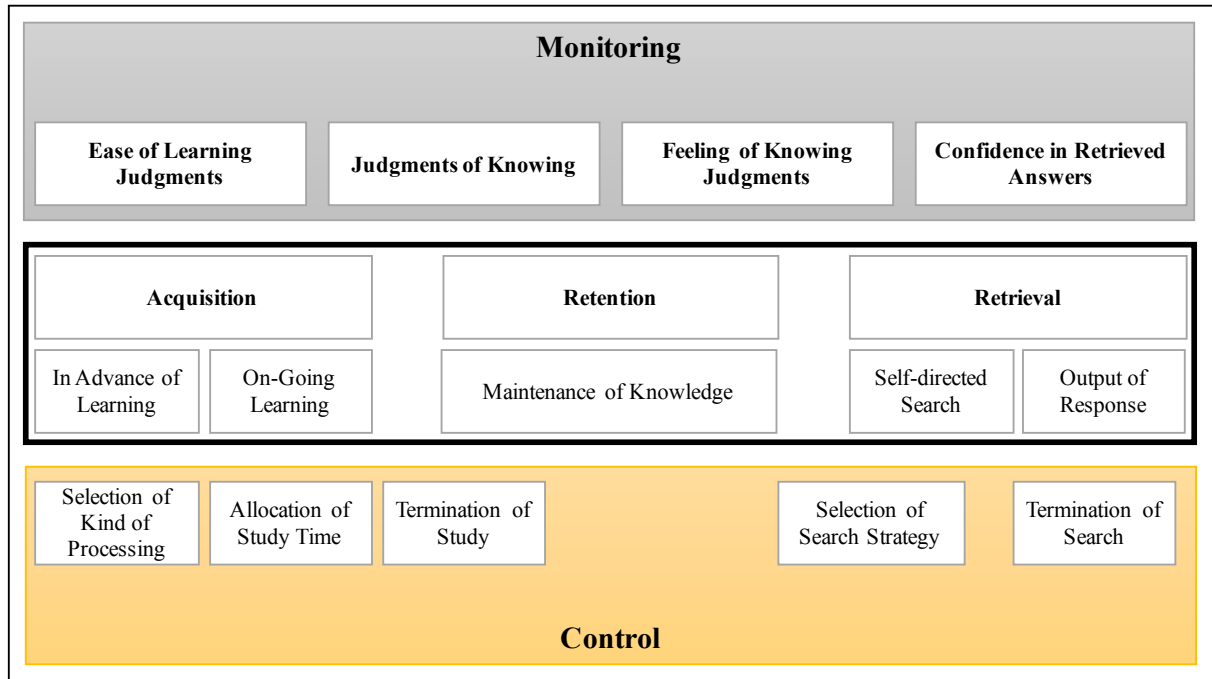


Abbildung 4: Kontroll- und Monitoringprozesse nach Nelson & Narens (1990)

Monitoring- und Kontrollprozesse sind von zentraler Bedeutung für das Modell von Nelson & Narens (1990). An unterschiedlichen Stellen des Lernprozesses werden verschiedene Prozesse des Monitorings und der Kontrolle vollzogen.

Vor Beginn des Lernprozesses nehmen Lernende nach dem Modell von Nelson & Narens (1990) *Ease of Learning*-Urteile vor (EoL). Hierbei geben Lernende bevor sie beginnen, Inhalte zu lernen, eine Einschätzung darüber ab, wie leicht/schwierig ihnen das Lernen dieser Inhalte fallen wird (Dunlosky & Metcalfe, 2009; Leonesio & Nelson, 1990; Zechmeister & Bennett, 1991). Sie gehören damit zu den prospektiven Selbsteinschätzungen (Nelson & Narens, 1990). Auf Basis dieser Selbsteinschätzung werden Entscheidungen für den weiteren Lernprozess getroffen (*selection of kind of processing*) und die Zeitaufteilung für das Lernen der Inhalte (*allocation of study time*) vorgenommen.

In der Phase des Wissenserwerbs werden Urteile über das eigene Wissen bzw. das eigene Lernen vorgenommen. In der Literatur werden diese als *Judgements of Learning* (JoL) oder auch *Judgements of Knowing* (JoK) bezeichnet. Sie beinhalten eine Vorhersage über die Performanz in einer zukünftigen Testung zu aktuell gelernten Inhalten (Nelson & Narens, 1990). Koriart (1997) definiert *Judgements of Learning* als Urteile, die am Ende einer Lerneinheit getroffen

werden und die Wahrscheinlichkeit, die gelernten Informationen wiedergeben zu können, umschreiben. Experimentell wird in diesem Kontext üblicherweise dergestalt vorgegangen, dass sich Lernende zunächst eine Liste von Wortpaaren einprägen. Nach der Lerneinheit werden sie um die Selbsteinschätzung gebeten, wie viele Wortpaare sie richtig benennen können (Koriat, 1997).

Während der Erarbeitungsphase der Lerninhalte, kommen darüber hinaus sogenannte *Feeling of Knowing*-Urteile zum Tragen. Diese Urteile beziehen sich auf die Einschätzung, inwiefern ein aktuell noch nicht wiedergebbbarer Lerninhalt in einer zukünftigen Testung wiedergegeben werden kann (Nelson & Narens, 1990, S. 130).

Zu den retrospektiven Selbsteinschätzungen gehören die *Confidence Judgements* (CJ) nach der Testung zu den gelernten Inhalten. Mit diesen geben die Studierenden die Wahrscheinlichkeit an, dass die von ihnen gegebene Antwort auf eine Aufgabenstellung die richtige ist (Dunlosky & Metcalfe, 2009).

Auf Basis dieser Ausdifferenzierung des Monitorings lassen sich vier potentielle Initiierungszeitpunkte für eine Selbsteinschätzung definieren, die im weiteren Fortgang dieser Arbeit zur Kategorisierung von Selbsteinschätzungen genutzt werden. In einem ersten Schritt soll nun jedoch eruiert werden, inwiefern Lernende in der Lage sind, akkurate Selbsteinschätzungen vorzunehmen und welche Effekte mit dem Einsatz von Instrumenten der Selbsteinschätzung verbunden sind.

2.2 Die Genauigkeit der Selbsteinschätzungen von Lernenden

Die Fähigkeit der Lernenden zur Einschätzung des eigenen Wissensstandes ist Gegenstand einer Reihe von Metastudien und Reviews (u. a. Blanch-Hartigan, 2011; Brown & Harris, 2013; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989). Die Studien berücksichtigen dabei unterschiedliche Kontexte, in denen Selbsteinschätzungen erfolgen.

Brown & Harris (2013) betrachten empirische Studien zum Einsatz von Self-Assessments in der Schule innerhalb des K-12-Sektors¹. Auf Basis von 84 ausgewählten Studien wird u. a. der Frage nach den Zusammenhängen zwischen der Selbsteinschätzung und der Performanz sowie der Fähigkeit zur Selbstregulation nachgegangen. Ferner ermitteln Brown & Harris (2013) Einflussfaktoren, die die Fähigkeit zur Selbsteinschätzung bedingen.

Die weiteren angeführten Metastudien konzentrieren sich auf Self-Assessments im universitären Kontext. Einen ersten umfassenden Überblick über die Fähigkeit zur Selbsteinschätzung von Studierenden geben Falchikov & Boud (1989) in ihrem Review. Sie analysieren 48 quantitative Studien mit dem Ziel, die Akkuratheit der Selbsteinschätzung von Studierenden sowie deren Bedingungsfaktoren zu ermitteln. Als Bedingungsfaktoren werden die Qualität des Studiendesigns², der Lernstand der Lernenden, die Form des implementierten Self-Assessments, die thematische Einheit, in der das Self-Assessment eingesetzt wird, und das Instrument, welches der Selbsteinschätzung zugrunde liegt, berücksichtigt.

Dochy et al. (1999) analysieren 63 Studien, die den Einsatz von Self-, Peer- und Co-Assessments³ in der universitären Bildung zum Gegenstand haben, mit dem Ziel, Handlungsempfehlungen für den Einsatz von Self-Assessments in der Praxis zu entwickeln.

Das Vorgehen zur Bestimmung der Akkuratheit der Selbsteinschätzung differiert in den genannten drei Metastudien wenig. Die Fähigkeit zur akkuraten Selbsteinschätzung wird im schulischen Kontext in der Regel über Vergleiche der Selbsteinschätzung mit den Einschätzungen der Lehrpersonen, dem aktuellen Abschneiden der Lernenden bei objektiven Tests und

¹ Der K-12-Sektor bezeichnet in den USA die schulische Bildung vom Kindergarten bis zum zwölften Schuljahr.

² Zur Überprüfung der Qualität des Studiendesigns wurden verschiedene Indikatoren herangezogen wie bspw. die Angemessenheit der eingesetzten Aufgaben und Kriterien zur Selbsteinschätzung, die Reliabilität und Validität der Instrumente sowie die Nachvollziehbarkeit der berichteten Informationen. Auf dieser Basis wurden die Studien in qualitativ hochwertige und weniger hochwertige Studien eingeteilt (Falchikov & Boud, 1989, S. 416).

³ Nach Dochy et al. (1999) unterscheiden sich Co-Assessments von Self- und Peer-Assessments dahingehend, dass sie den Lernenden zwar die Möglichkeit geben, am Beurteilungsprozess zu partizipieren, die Kontrolle hierüber jedoch auf Seiten des Dozenten verbleibt. Demnach eignen sich diese eher für summative Assessments, während Self- und Peer-Assessments formative Assessments ermöglichen.

rubricbasierten Einschätzungen durch die Lehrpersonen ermittelt. Wie akkurat die Fremdeinschätzungen der Lehrpersonen hierbei sind, bleibt jedoch unberücksichtigt⁴. Analog dazu werden im universitären Kontext oftmals die Einschätzungen der Dozierenden herangezogen. Die entsprechenden Korrelationen zwischen Selbst- und Fremdeinschätzung sind nach Brown & Harris (2013, S. 384) durchgängig positiv ($r = .20$ bis $.80$). Jedoch berichten nur wenige Studien eine Korrelation von $r > .60$, sodass die durchschnittliche Korrelation eher als moderat einzustufen sein dürfte. Zu ähnlichen Werten gelangen auch Falchikov & Boud (1989). Zur Bestimmung der Akkuratheit der Selbsteinschätzungen ermitteln sie Effektstärken und Korrelationen. Die Effektstärken variieren über die 48 Studien zwischen $d = -.62$ und $d = 1.42$, wobei die mittlere Effektstärke $d = .47$ beträgt (Falchikov & Boud, 1989, S. 419). Falchikov & Boud (1989) verweisen darauf, dass die Autoren der betrachteten Studien mehrheitlich zu dem Ergebnis gelangen, dass Studierende ihren Wissensstand tendenziell überschätzen. Die von ihnen referierten Korrelationen zwischen Selbsteinschätzungen der Lernenden und Einschätzungen der Lehrenden weisen Werte zwischen $r = -.05$ und $r = .82$ auf. Die mittlere Korrelation über alle Studien hinweg beträgt $r = .39$. Darüber hinaus werden auch Studien betrachtet, in denen die prozentuale Übereinstimmung von Selbsteinschätzung und Einschätzung durch den Lehrenden zur Bestimmung der Akkuratheit herangezogen wurde. Die im Rahmen der Studien ermittelten Übereinstimmungsprozente liegen nach Falchikov & Boud (1989, S. 420) zwischen 33 % und 99 %, wobei die mittlere Übereinstimmung 64 % beträgt. Dochy et al. (1999) bestätigen in ihren referierten Studien die in Teilen moderat ausgeprägten Zusammenhänge. Zahlenwerte werden jedoch nur aus der Studie von Longhurst & Norton (1997) genannt ($r = .43$).

Auffällig sind die recht hohen Spannweiten zwischen den in den einzelnen Studien referierten Zusammenhängen zwischen Selbst- und Fremdeinschätzung. Aus diesem Grund untersuchen die Autoren der Metastudien Faktoren, die diese Varianz erklären könnten.

Nach Brown & Harris (2013) wird die Akkuratheit der Selbsteinschätzung vom Alter und den schulischen Erfahrungen der Lernenden, der Aufgabenstellung, die der Selbsteinschätzung zugrunde liegt, und dem Leistungsniveau der Lernenden beeinflusst. Ältere und somit erfahrenere Lernende (Sekundarbereich) weisen bessere Fähigkeiten zur akkuraten Selbsteinschätzung auf als jüngere. Unklar ist jedoch, inwiefern die verbesserte Akkuratheit mit zunehmendem Alter

⁴ Die diagnostische Kompetenz von Lehrpersonen ist Gegenstand einiger Metastudien. Dabei zeigt sich, dass Lehrpersonen zur Einschätzung der Schülerleistungen weitgehend in der Lage sind. Die Korrelationswerte zwischen Lehrereinschätzungen und Testleistungen der Lernenden variieren jedoch zwischen den Studien (Anders, Kunter, Brunner, Krauss, & Baumert, 2010; Hoge & Coladarci, 1989; Südkamp, Kaiser, & Möller, 2012). Einflussfaktoren bzgl. der Genauigkeit der Lehreinschätzung sind u. a. die Domäne, die Skalierung der Beurteilungsinstrumente und die Spezifität des Lehrerurteils (Hoge & Coladarci, 1989).

Folge eines allgemeinen Entwicklungsprozesses ist oder auf die zunehmenden schulischen Erfahrungen zurückzuführen ist. Hinsichtlich der Performanz der Lernenden ist zu beobachten, dass eine höhere Performanz grundsätzlich mit einer höheren Akkuratheit der Selbsteinschätzung einhergeht. Überdies sind die Korrelationswerte höher, wenn eher einfache und hinsichtlich der Einschätzung der Richtigkeit gut zu prognostizierende Aufgaben (z. B. geschlossene Aufgabenformate) eingesetzt werden. In diesem Zusammenhang ist auch bedeutsam, wie konkret die Anforderungskriterien zur Einschätzung formuliert sind. Je spezifischer die Kriterien sind, desto akkurater ist die Selbsteinschätzung der Lernenden. Nicht näher beleuchtet wird, inwiefern hierfür auch die (In-)Akkuratheit der Beurteilung durch die Lehrpersonen eine Rolle spielt.

Dochy et al. (1999) gelangen hinsichtlich der Akkuratheit der Selbsteinschätzung und deren Einflussfaktoren zu ähnlichen Ergebnissen wie Brown & Harris (2013). Die Akkuratheit der Selbsteinschätzung ist demnach moderat und hängt vom Leistungsniveau der Lernenden („high-performer“ weisen akkuratere Werte bei der Selbsteinschätzung auf) ab. Außerdem stellen sie fest, dass Studierende in höheren Semestern in ihren Selbsteinschätzungen akkurater sind. Neben reinen Self-Assessments betrachten sie auch den kombinierten Einsatz von Self- und Peer-Assessments. Die Übereinstimmung zwischen Selbsteinschätzungen und Einschätzungen durch die Peers ist sehr hoch (Korrelationen zwischen den Einschätzungen liegen bei $r = .93$ bis $r = .99$). Eine Tendenz zur Über-/Unterschätzung lässt sich nach Dochy et al. (1999) nicht feststellen.

Hinsichtlich der Einflussfaktoren auf die Güte der Selbsteinschätzung identifizieren Falchikov & Boud (1989, S. 422) zunächst einen Effekt des Studiendesigns. In Studien mit einem qualitativ besseren Design ist die relative Übereinstimmung zwischen dem Self-Assessment des Lernenden und der Einschätzung/Beurteilung des Lehrenden höher als in Studien mit schwächerem Design (69.62 % vs. 41.83 %). Allerdings deckt sich dieser Befund nicht mit Ergebnissen aus Studien, in denen die Korrelationen zwischen Selbst- und Fremdeinschätzung zur Bestimmung der Akkuratheit der Selbsteinschätzung ermittelt wurden. Diese sind in Studien mit schwächerem Design höher ausgeprägt als in Studien mit qualitativ besserem Design. Allerdings bewerten Falchikov & Boud (1989) nur wenige Studien als qualitativ schwache Studien ($n = 5$) im Vergleich zu qualitativ hochwertigen Studien ($n = 25$).

Die ermittelten Ergebnisse zum Einfluss des Studienfortschritts auf die Akkuratheit der Selbsteinschätzung sind ebenso different. Studien, in denen die prozentuale Übereinstimmung zwischen Selbst- und Fremdeinschätzung bestimmt worden ist, zeigen, dass die Akkuratheit der

Selbsteinschätzung bei weiter fortgeschrittenen Studierenden höher ist als bei Studierenden, die sich noch am Anfang des Studiums befinden. In Studien die wiederum Korrelationen zwischen Selbsteinschätzung und Studienfortschritt bestimmten, fällt die Fähigkeit zur Selbsteinschätzung bei Studierenden im ersten Studienjahr präziser aus als bei Studierenden, die sich in höheren Semestern befinden ($r = .54$ vs. $r = .33$).

Geringe Unterschiede wurden in der Akkuratheit der Selbsteinschätzung bezüglich der eingesetzten Instrumente (z. B. Einsatz von Rubrics/Skalen zur Selbsteinschätzung) registriert. Tendenziell führt eine stärkere Beteiligung der Lernenden an der Entwicklung der Kriterien, die zur Selbst- und Fremdeinschätzung verwendet werden sollen, zu einer höheren Akkuratheit der Selbsteinschätzung. Darüber hinaus sind die Selbsteinschätzungen in naturwissenschaftlichen Fächern genauer als in sozialwissenschaftlichen Disziplinen.

Die Betrachtung der maßgeblichen Reviews und Metastudien zur Akkuratheit der Selbsteinschätzungen offenbart, dass Lernende bedingt in der Lage sind, ihren Wissensstand genau einzuschätzen. Zur Bestimmung der Akkuratheit der Selbsteinschätzung wurden Korrelationen und Übereinstimmungsprozente ermittelt sowie Mittelwertvergleiche durchgeführt. Die Ergebnisse variieren über die Studien hinweg relativ stark. Als Einflussfaktoren werden sowohl lernerspezifische Komponenten (z. B. allgemeine Fähigkeiten der Lernenden, Alter und Erfahrung) aber auch studienspezifische Komponenten (z. B. Fachrichtung, eingesetzte Instrumente des Self-Assessments, Kurslevel) angeführt. Abbildung 5 gibt eine Übersicht über die identifizierten Faktoren.

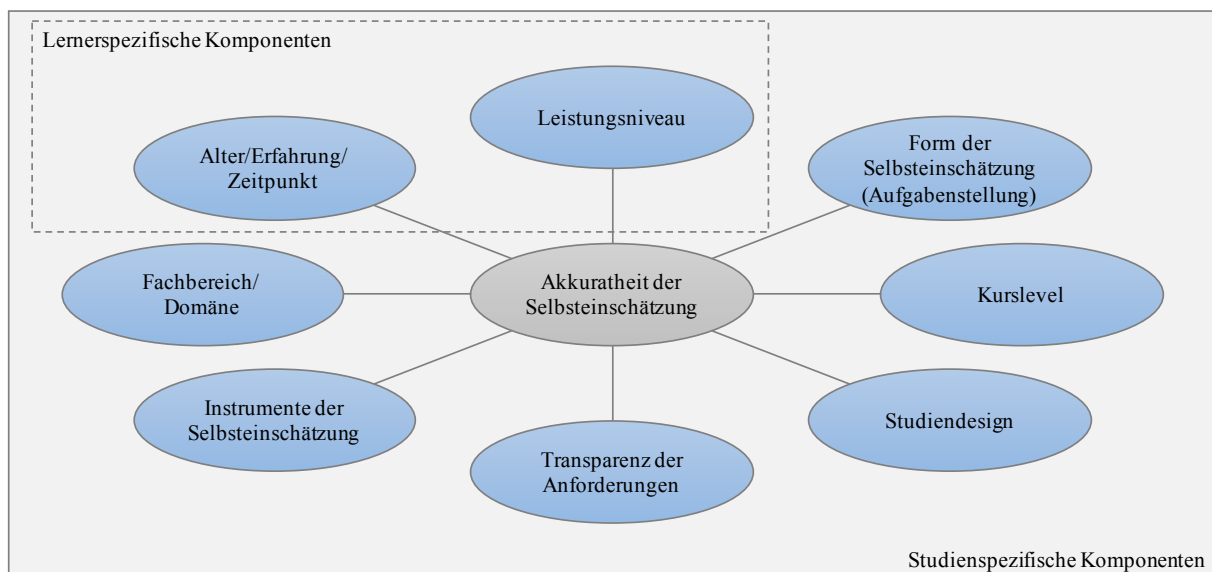


Abbildung 5: Einflussfaktoren auf die Akkuratheit der Selbsteinschätzung

2.3 Effekte der Initiierung von Selbsteinschätzungen

Neben der Fähigkeit zur Selbsteinschätzung wurden in den bereits zitierten Reviews und Metastudien in Teilen auch die Effekte des Einsatzes von Self-Assessments auf das Lernen untersucht.

Brown & Harris (2013) identifizieren 21 Studien, welche die Effekte des Einsatzes von Self-Assessments auf das Lernen zum Untersuchungsgegenstand haben. Der Median der ermittelten Effektstärken wird von Brown & Harris (2013, S. 381) mit $d = .40$ (Ross, Hogaboam-Gray, & Rolheiser, 2002) bzw. $d = .45^5$ (McDonald & Boud, 2003) angegeben. In der Aufstellung der Studien nach absteigender Effektstärke führen Brown & Harris (2013) auch die in den verschiedenen Studien eingesetzten Formen von Selbsteinschätzungen an (z. B. Selbsteinschätzung auf der Basis von Rubrics). Auffällig dabei ist, dass in den referierten Studien keine Form der Selbsteinschätzung durchgängig hohe Effektstärken in Bezug auf die Performanz aufweist. So werden bspw. für Selbsteinschätzungen, die mit Hilfe von Rubrics vorgenommen werden, in einigen Studien hohe Effektstärken festgestellt (z. B. Andrade, Du, & Wang, 2008; Andrade, Du, & Mycek, 2010; Sadler & Good, 2006), während in anderen Studien nur geringe Effektstärken berichtet werden (Andrade & Boulay, 2003; Ross, Hogaboam-Gray, & Rolheiser, 1999). Demnach werden Self-Assessments insgesamt zwar positive Effekte auf das Lernergebnis zugeschrieben, welche Form der Selbsteinschätzung sich als besonders vorteilhaft erweist, ist jedoch nicht abschließend geklärt. Unklar ist ebenfalls, inwieweit positive Effekte auch bzgl. der Realisierung selbstregulierten Lernens bestehen. Gleichwohl Brown & Harris (2013) tendenziell positive Effekte auf Motivation und Selbstregulation der Lernenden konstatieren, sind die Studienergebnisse nicht eindeutig.

Darüber hinaus bestätigen Dochy et al. (1999) die positiven Effekte auf das Lernen, die dem Einsatz von Self-Assessments zugeschrieben werden. Sie referieren Studienergebnisse (z. B. Hassmén, Sams, & Hunt, 1996), nach denen Lernende nach dem Einsatz von Instrumenten des Self-Assessments mehr korrekte Antworten im Rahmen von Tests geben als Lernende, die solche Instrumente nicht erhielten.

In jüngster Vergangenheit durchgeführte Studien bestätigen die Ergebnisse zu den Effekten des Einsatzes von Self-Assessments. Andrade, Du, & Mycek (2010) untersuchen für den Englischunterricht den Zusammenhang zwischen dem Einsatz von Rubrics als Instrument zur Selbsteinschätzung des Wissensstandes und den Fertigkeiten von Lernenden der fünften bis siebten

⁵ Brown & Harris (2013) referieren 22 Effektstärken von $d = -.04$ bis $d = 1.62$. Zu einer der 21 herangezogenen Studien konnten zwei Effektstärken ermittelt werden.

Klasse, Essays zu verfassen. Ferner war es Ziel der Studie, geschlechtsspezifische Unterschiede bezüglich des Self-Assessments zu kontrollieren. Hierzu erhielten Lernende der Treatment-Gruppe einen Rubricbogen mit sieben inhaltlichen Kriterien bzgl. des Schreibens von Texten, die auf einer vierstufigen Skala eingeschätzt werden mussten. Die Lernenden wurden nach Diskussion und Anwendung der Rubrics an einem Beispiel angehalten, diesen zur Selbsteinschätzung zu nutzen und Kriterien in den Rubrics sowie deren Entsprechung in den eigenen Texten zu markieren. Die Kontrollgruppe schätzte die Güte der erstellten Texte ohne Rubrics selbst ein. Es zeigte sich, dass die Lernenden der Treatment-Gruppe signifikant bessere Texte als die Lernenden der Kontrollgruppe verfassten. Das Geschlecht der Lernenden hatte keinen Einfluss auf den Effekt.

Panadero, Tapia, & Huertas (2012) untersuchen die Effekte des Einsatzes zweier Instrumente zur Selbsteinschätzung (Rubrics & Scripts⁶) auf Selbstregulation, Selbstwirksamkeit und Lernergebnis. Hierbei zeigt sich, dass Scripts bezüglich der Selbstregulation einen größeren unterstützenden Effekt ausüben als der Einsatz von Rubrics. Grundsätzlich kommen Panadero et al. (2012) zum Ergebnis, dass mit dem Einsatz von Instrumenten der Selbsteinschätzung positive Effekte auf das Lernen verbunden sind.

Auch Barney, Khurum, Petersen, Unterkalmsteiner, & Jabangwe (2012) untersuchen den Effekt des Einsatzes von Rubrics zur Selbsteinschätzung auf das Lernen. Hierzu ließen sie Aufgabebearbeitungen von Ingenieurstudenten anhand von sieben Kriterien auf einer fünfstufigen Skala einschätzen. Barney et al. (2012) stellen fest, dass die Lernenden durchaus zur akkuraten Selbsteinschätzung in der Lage waren (45 % Übereinstimmung mit der Einschätzung der Dozenten). Darüber hinaus verbesserten sich die Lernergebnisse signifikant, sodass ebenfalls ein positiver Effekt auf das Lernen konstatiert werden kann.

⁶ Rubrics enthalten nach Panadero et al. (2012) Angaben über relevante Beurteilungskriterien, eine Skala zur Einschätzung der Zielerreichungsgrade sowie eine Beschreibung eines jeden Zielerreichungsgrades. Scripts sind aufgabenbegleitende Fragestellungen, die sich an den Lösungsschritten eines Experten bei der Bearbeitung einer Aufgabe orientieren.

3 Empirische Befunde zum Zusammenhang zwischen Akkuratheit der Selbsteinschätzung, selbstreguliertem Lernen und der Performanz

3.1 Studienrecherche

Die Darstellung bisheriger Studienergebnisse zur Akkuratheit der Selbsteinschätzung und der Performanz zeigte, dass die Lernenden bedingt in der Lage sind, ihre Performanz akkurat einzuschätzen (Brown & Harris, 2013; Falchikov & Boud, 1989), aber grundsätzlich positive Effekte mit der Initiierung von Selbsteinschätzungen verbunden werden (Panadero & Romero, 2014; Panadero et al., 2012). Somit scheint die Fragestellung nach der Bedeutung möglichst akkurater Selbsteinschätzungen für die Performanz noch nicht hinreichend geklärt zu sein. Es ist denkbar, dass allein die Initiierung einer Selbsteinschätzung unabhängig von ihrer Akkuratheit bereits einen Effekt auf das Lernen hat.

Um dieser Frage nachzugehen, werden im Folgenden Studienergebnisse präsentiert, die die Auswirkungen einer (in-)akkuraten Selbsteinschätzung auf den weiteren Lernprozess zum Gegenstand haben. Es werden demnach im besonderen Maße Studien berücksichtigt, die sowohl die Akkuratheit der Selbsteinschätzung als auch die Effekte der Selbsteinschätzung beinhalten. Zur Literaturrecherche wurden die Datenbanken von EBSCO (hierunter waren insbesondere „Academic Search Premier“, „Academic Search Alumni Edition“ und „PsycINFO“ von Bedeutung) und ERIC genutzt. In einem ersten Schritt wurden mit den Schlagworten „self-assessment accuracy“, „self-evaluation accuracy“, „self-judgement accuracy“, „self-rating accuracy“ und „self-grading accuracy“ Studien identifiziert, die explizit die Akkuratheit der Selbsteinschätzung betrachten. Die Abstracts wurden dahingehend analysiert, inwieweit auch der Einfluss der Akkuratheit der Selbsteinschätzung auf das Lernen untersucht wurde.

Die Recherche mit den genannten Schlagworten ergab insgesamt 233 Studien (beinhaltet auch doppelte Nennungen in den Datenbanken ERIC und EBSCO). Von den 233 Studien wurden nach Sichtung der jeweiligen Abstracts 28 Studien einer genaueren Lektüre unterzogen. Es zeigte sich jedoch, dass nicht alle Studien die zugrunde gelegten Kriterien (Prüfung von Zusammenhängen zwischen Akkuratheit der Selbsteinschätzung und der Performanz) erfüllten.

Entsprechend nicht berücksichtigt wurden Studien, die:

- Selbsteinschätzungen außerhalb des Kontextes Lehren und Lernen zum Gegenstand hatten.
- nicht explizit die Akkuratheit der Selbsteinschätzung in Betracht zogen (z. B. Andrade et al., 2010; Andrade, Wang, Du, & Akawi, 2009; Panadero et al., 2012).

- keine Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung und Performanz oder Elementen des selbstregulierten Lernens aufwiesen. In einigen Studien wurden zwar Selbsteinschätzung und Performanz der Probanden getestet, jedoch diente dies nur zur Bestimmung der Akkuratheit der Selbsteinschätzung. Weiterführende Untersuchungen zu den Zusammenhängen zwischen der Fähigkeit zur Selbsteinschätzung und der Performanz wurden nicht vorgenommen (z. B. Gillström & Rönnerberg, 1995; Thiede & Dunlosky, 1999).
- lediglich anderweitig durchgeführte und publizierte Studien repetieren und gegenüberstellen (z. B. Butler, 2011; Dunlosky & Hertzog, 2005).

Die recherchierten Studien (N = 19) wurden hinsichtlich der Kriterien „theoretischer Bezugsrahmen“, „Fragestellungen/Hypothesen“, „Design“ und „Ergebnisse“ in tabellarischer Form zusammengefasst (s. Anhang A1). Im Allgemeinen werden die Studien dahingehend betrachtet, inwiefern sie die Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung, der Selbstregulation und der Performanz abbilden. Diese Zusammenhänge können in Teilprozesse, die sich an den verschiedenen Initiierungszeitpunkten von Self-Assessments orientieren (s. Kapitel 2.1), ausdifferenziert werden.

Grundlegend lassen sich vier Initiierungszeitpunkte für ein Self-Assessment unterscheiden (s. Abbildung 6).

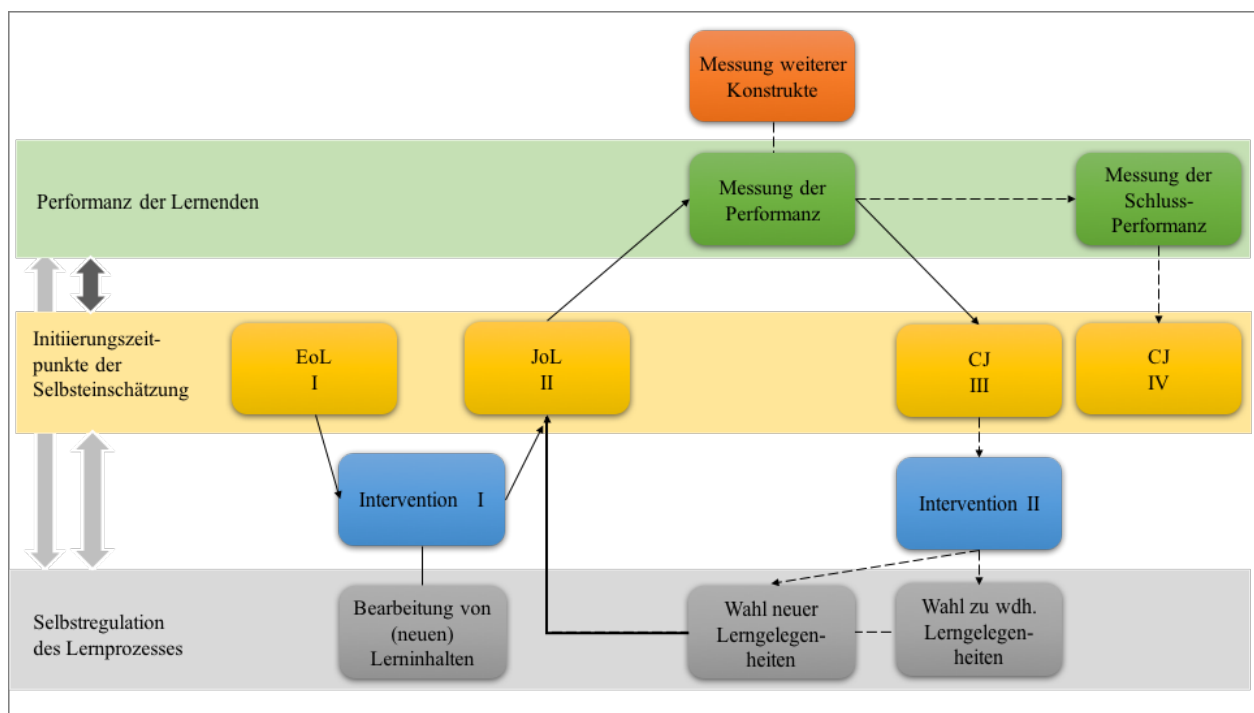


Abbildung 6: Initiierungszeitpunkte der Selbsteinschätzung

Eine erste Selbsteinschätzung kann vor dem Lernprozess i. S. eines *EoL*-Urteils erfolgen (Initiierungszeitpunkt I). Nach dieser Einschätzung wird die erste Intervention terminiert. An dieser Stelle bearbeiten die Lernenden (neue) Lerninhalte, die in der Regel vorgegeben werden. Vor und nach der Messung des in der ersten Intervention Gelernten, können die Lernenden ihre Performanz i. S. eines *Judgements of Learning* (Initiierungszeitpunkt II) oder *Confidence Judgements* (Initiierungszeitpunkt III) einschätzen. Ein analoges Vorgehen ergibt sich, wenn statt der Performanz weitere, im Kontext des Lernens bedeutsame Konstrukte auf ihren Zusammenhang mit der Akkuratheit der Selbsteinschätzung geprüft werden. Im Folgenden kann der Prozess, bestehend aus Lernphase, Selbsteinschätzung und Testung des Gelernten im Rahmen eines selbstregulierten Lernens wiederholt werden. Abschließend erfolgt eine finale Messung der Performanz, die von einer weiteren Selbsteinschätzung (Initiierungszeitpunkt IV) gefolgt sein kann.

Studien, in denen alle Initiierungszeitpunkte für eine Selbsteinschätzung berücksichtigt werden, lassen sich nicht finden. Eine Kategorisierung der Studienergebnisse in die folgenden drei Bereiche ist jedoch möglich:

1) Effekte (in-)akkuratere Selbsteinschätzungen auf die Selbstregulation und die Performanz

Diese Kategorie umfasst Studien, in denen Zusammenhänge von Selbsteinschätzung, Selbstregulation des Lernens und Performanz berücksichtigt werden. Das Vorgehen in diesen Studien ist dergestalt, dass Probanden nach einer Lernphase eine Selbsteinschätzung bzgl. ihrer (erwarteten) Performanz vornehmen. Anschließend wird untersucht, wie die Akkuratheit der Selbsteinschätzung zunächst auf die Regulation des weiteren Lernprozesses (z. B. Wahl neuer Lerngelegenheiten) wirkt und schließlich mit der Performanz in einem abschließenden Test zusammenhängt. In diesen Studien erfolgt nach einer Selbsteinschätzung zum Initiierungszeitpunkt II oder III mindestens eine weitere selbstregulierte Lernphase oder eine Einschätzung des eigenen selbstregulativen Verhaltens, bevor die abschließende Performanztestung vorgenommen wird.

2) Effekte (in-)akkuratere Selbsteinschätzungen auf die Performanz

Die Studien der zweiten Kategorie berücksichtigen den selbstregulierten Lernprozess nach einer Selbsteinschätzung (Initiierungszeitpunkt II oder III) im Gegensatz zu Studien der ersten Kategorie nicht, bilden aber die Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der Performanz ab. Sie unterscheiden sich somit von einer Vielzahl an Studien, die lediglich Effekte der Initiierung von Selbsteinschätzungen unabhängig von der

Akkuratheit ebendieser eruieren. Herangezogen werden Studienergebnisse, die unabhängig vom Initiierungszeitpunkt der Selbsteinschätzung Rückschlüsse auf den Zusammenhang von Akkuratheit der Selbsteinschätzung und Performanz erlauben.

3) Effekte (in-)akkurater Selbsteinschätzungen auf weitere lernrelevante Faktoren

In der dritten Kategorie sind jene Studien enthalten, die nicht explizit auf die Performanz, sondern auf weitere lernbedeutsame Konstrukte (z. B. Motivation, Emotionen) abzielen. Gemeinsam haben diese Studien mit jenen der ersten beiden Kategorien, dass sie Zusammenhänge mit der Akkuratheit der Selbsteinschätzung (Initiierungszeitpunkte I bis IV) prüfen.

3.2 Studienergebnisse

3.2.1 Effekte (in-)akkuratere Selbsteinschätzungen auf Selbstregulation und Performanz

In einem ersten Schritt werden in der Folge Studienergebnisse dargestellt, die die Auswirkungen von (in-)akkuraten Selbsteinschätzungen auf die Selbstregulation zum Gegenstand der Untersuchung haben und dabei weitergehend Aussagen über die Performanz von Lernenden mit (in-)akkuraten Selbsteinschätzungen ermöglichen. Bislang wurde in wenigen Studien untersucht, welche konkreten Auswirkungen eine (in-)akurate Selbsteinschätzung des Wissensstandes auf die Wahrnehmung weiterer Lerngelegenheiten seitens der Lernenden und der Performanz hat. Tabelle 1 gibt einen Überblick über jene Studien, die den Dreischritt von Selbsteinschätzung, Selbstregulation und Performanz untersucht haben.

Studie	Selbsteinschätzung	Initiierungszeitpunkt	Selbstregulation	Messung der Performanz
Dunlosky & Rawson (2012)	Confidence Judgements	III	Auswahl wiederholt zu lernender Inhalte (Definitionen aus Texten)	Abfrage von gelernten Definitionen
Thiede, Anderson, & Theriault (2003)	Ease of Learning Judgements Judgements of Learning	I & II	Auswahl wiederholt zu lernender Inhalte (Texte)	Beantwortung von Fragen zu gelernten Texten
van Loon, de Bruin, van Gog, & van Merriënboer (2013)	Judgements of Learning Self-Score-Judgements (~ Confidence Judgements)	II & IV	Auswahl wiederholt zu lernender Konzepte	Abfrage von gelernten Konzepten
Kostons, van Gog, & Paas (2010)	Judgements of Learning zu verschiedenen Kriterien	II	Auswahl neuer Lernaufgaben mit unterschiedlichen Anspruchsniveaus	MC-Aufgaben zu Genetik
Kostons, van Gog, & Paas (2012)	Judgements of Learning zu verschiedenen Kriterien	II	Auswahl neuer Lernaufgaben mit unterschiedlichen Anspruchsniveaus	Aufgaben zu Genetik
Sitzmann & Johnson (2012)	Judgements of Learning	II	Weitere Teilnahme am Schulungsprogramm	MC-Aufgaben zu MS Excel
Panadero & Romero (2014)	Judgements of Learning	II	Einsatz von Skalen zur Selbstregulation (Emotion and Motivation Self-Regulation Questionnaire & Self-Regulation Open Questions)	Entwicklung einer Concept Map

Tabelle 1: Überblick über Studien der ersten Kategorie

Dunlosky & Rawson (2012) analysierten in zwei durchgeführten Studien auf Basis der *Accuracy Influences Memory-Hypothese*⁷, nach der akkurates Monitoring zu effektiverem Lernen und letztlich besseren Lernergebnissen führen soll, die Zusammenhänge zwischen absoluter Akkuratheit⁸ der Selbsteinschätzung, der Wahrnehmung von nachgelagerten Lernmöglichkeiten und der Performanz in einem abschließenden Test. Für Dunlosky & Rawson (2012) bestand das zentrale Ziel der Studie darin, den vermuteten Zusammenhang zwischen der absoluten Akkuratheit der Selbsteinschätzung und der Performanz in einem abschließenden Test zu prüfen. Lernende, die ihren Wissensstand unterschätzen, könnten von einem Effekt des Überlernens profitieren, wohingegen Lernende bei Überschätzung den Lernprozess möglicherweise zu früh beenden und schlechtere Leistungen im abschließenden Test aufweisen könnten.

Im Rahmen eines ersten Experimentes lernten die Probanden in einer computergestützten Lernumgebung (N = 48 Studierende der Psychologie) selbstgesteuert Definitionen aus Texten. Nach dem ersten Lernen sollten sie die Definitionen wiedergeben und die Richtigkeit ihrer Antwort auf einer dreistufigen Skala einschätzen (Initiierungszeitpunkt III). Die Treatment-Gruppe erhielt bei der Selbsteinschätzung Unterstützung in Form von Prompts (Schlüsselemente der Definitionen), die parallel zur gegebenen Antwort gezeigt wurden. Die Kontrollgruppe erhielt diese Prompts nicht. Nach der Selbsteinschätzung wurde allen Probanden die korrekte Antwort gezeigt und sie bekamen in der Folge die Möglichkeit, die Aufgaben zeitlich selbstbestimmt wiederholt zu lernen. Sobald die Erinnerung einer Definition drei Mal als korrekt eingeschätzt worden ist, wurde diese von der eingesetzten Software aus der Liste der potentiell wiederholt zu bearbeitenden Items genommen. Zwei Tage nach der Lerneinheit erfolgte die abschließende Performanztestung, bei der jedoch keine weitere Selbsteinschätzung gefordert war.

Es zeigte sich, dass die Probanden der Treatment-Gruppe akkuratere Selbsteinschätzungen vornahmen als Lernende der Kontrollgruppe, die sich tendenziell stärker überschätzten. Dabei unterschieden sich die Lernenden beider Versuchsgruppen nicht im Zeitaufwand, mit dem sie die Definitionen (wiederholt) lernten. Allerdings stellten Dunlosky & Rawson (2012) fest, dass Lernende mit akkurater Selbsteinschätzung eher nicht gewusste Items zum wiederholten Lernen auswählten als Lernende, die ihre Testleistungen überschätzten. Erwartungsgemäß zeigte sich letztlich, dass die in ihren Selbsteinschätzungen akkurateren Lernenden in der Treatment-

⁷ Der Hypothese liegt die Annahme zugrunde, dass Lernende das Monitoring des eigenen Lernprozesses auch effektiv dazu nutzen, das Lernen zu kontrollieren. Weiterhin wird angenommen, dass Lernende sich im Umgang mit den Erkenntnissen aus dem Monitoring unterscheiden (Dunlosky & Rawson, 2012).

⁸ In der Literatur wird zwischen absoluter und relativer Akkuratheit der Selbsteinschätzung unterschieden. Die absolute Akkuratheit beschreibt die Passung zwischen vorhergesagter Performanz und der tatsächlichen Realisierung. Die relative Akkuratheit beschreibt die Korrelationen in beiden Richtungen zwischen vorhergesagter Performanz und tatsächlicher Realisierung (Maki, 1998).

Gruppe signifikant bessere Ergebnisse im Performanztest erzielten als jene in der Kontrollgruppe. Dunlosky & Rawson (2012) merken an, dass das Treatment nicht nur die Akkuratheit der Selbsteinschätzung verbesserte, sondern durch das Zeigen der Stichworte zur richtigen Lösung den Lernenden der Treatment-Gruppe ein Feedback gegeben wurde, was die beobachteten Unterschiede zur Kontrollgruppe bedingen könnte. Diesem Umstand trugen sie im Rahmen eines zweiten Experiments Rechnung, indem nun alle Probanden dem gleichen Setting wie die Lernenden aus der ersten Studie in der Treatment-Gruppe ausgesetzt wurden. Bei sonst gleichem Versuchsablauf bestätigten sich die Ergebnisse der ersten Studie. Um die Performanz als Funktion der Akkuratheit der Selbsteinschätzung zu prüfen, wurden die Lernenden in fünf Gruppen, abhängig vom Ausmaß der Überschätzung eingeteilt. Lernende, die in der selbstregulierten Lernphase mehr als 50 % ihrer Antworten irrtümlicherweise als korrekt eingestuft haben, weisen im abschließenden Performanztest weniger korrekte Antworten auf, als Lernende, die sich zuvor akkurat einschätzten (~ 30 % vs. ~ 90 % richtige Antworten). Je stärker die Überschätzung ausgeprägt ist, desto schlechter ist die Performanz der Lernenden im abschließenden Test.

Eine ähnliche Vorgehensweise wie Dunlosky & Rawson (2012) wählten Thiede, Anderson, & Theriault (2003) in ihrer Studie zu den Zusammenhängen zwischen akkuratem Monitoring, der Fähigkeit zur Selbstregulation und der Performanz. Auf Basis des Diskrepanz-Reduktions-Modells, wonach Lernende versuchen, wahrgenommene Diskrepanzen zwischen aktuellem Lernstand und Lernziel zu reduzieren, wird erwartet, dass eine akkuratere Selbsteinschätzung zu einer effektiveren Regulation des Lernens und somit zu besseren Lernergebnissen führt. Die Probanden (N = 66 Studierende der (Pädagogischen) Psychologie) wurden angehalten, nachdem sie ein EoL-Urteil auf einer siebenstufigen Skala gefällt haben (Initiierungszeitpunkt I), sechs Texte am Computer zu lesen und hierzu eigens Schlüsselbegriffe zu notieren. Dabei wurden sie randomisiert auf drei Gruppen aufgeteilt: Die erste Treatment-Gruppe notierte die Schlüsselbegriffe nach Bearbeitung aller Texte, während die zweite Treatment-Gruppe nach jedem gelesenen Text Notizen anfertigte. Die Kontrollgruppe notierte sich keine Stichworte. Im Anschluss sollten die Lernenden ihr Verständnis der Texte auf einer siebenstufigen Skala selbst einschätzen (Initiierungszeitpunkt II). Nach Bearbeitung eines Tests zu den gelernten Texten wählten die Studierenden eigenständig Texte zum wiederholten Lernen aus. Zuvor erhielten sie eine Rückmeldung über die Zahl der richtig beantworteten Fragen, jedoch keine Angaben darüber, welche Fragen korrekt beantwortet wurden. Nach der zweiten Lernphase erfolgte ein abschließender Test zum Verständnis der sechs Texte. Thiede et al. (2003) ermittelten für die erste Treatment-Gruppe eine signifikant höhere Akkuratheit der Selbsteinschätzung im

Vergleich zur zweiten Treatment-Gruppe und der Kontrollgruppe. Weiterhin konnten sie nachweisen, dass Lernende mit akkuraterer Selbsteinschätzung eher schlecht verstandene Texte wiederholten, was wiederum Auswirkungen auf das Abschneiden der Lernenden im finalen Test hatte. Waren die Testergebnisse vor dem Treatment noch für alle Gruppen vergleichbar, übertrafen Probanden der ersten Treatment-Gruppe jene aus der zweiten Treatment-Gruppe und der Kontrollgruppe im abschließenden Test signifikant. Diese Befunde sind konsistent mit jenen von Dunlosky & Rawson (2012). Die akkurate Selbsteinschätzung führte zu einer adäquateren Wahrnehmung weiterer Lerngelegenheiten und letztlich zur besseren Performanz im abschließenden Test.

Van Loon, de Bruin, van Gog, & van Merriënboer (2013) untersuchten, inwiefern die Aktivierung von inakkuratem Vorwissen einen Einfluss auf die Selbsteinschätzung und Selbstregulation des Lernenden besitzt. In zwei Studien wurden Primarschülerinnen und -schüler ($N = 103$ bzw. $N = 95$) dritter bis sechster Klassen angehalten, neue Konzepte zu lernen, das erworbene Wissen selbst einzuschätzen und einen Test zu bearbeiten, in dem die gelernten Konzepte erinnert werden mussten. In einem Pretest wurde zunächst eventuell bestehendes Vorwissen getestet. Für die weiteren Analysen unterschieden van Loon et al. (2013) daraufhin zwischen Lernenden ohne Vorwissen, Lernenden mit inakkuratem Vorwissen, tlw. akkuratem Vorwissen und akkuratem Vorwissen. Die Lernenden erhielten hierüber keine Rückmeldung. Insgesamt wurde nur zu einem Viertel der Items eine Antwort gegeben, die in der Mehrheit falsch war. Die Lernenden verfügten demnach über kein bzw. inakkurates Vorwissen. Anschließend lernten die Schülerinnen und Schüler selbstständig unter Bereitstellung von Lernmaterialien insgesamt 20 Konzepte. Nach der Lernphase erfolgte eine Selbsteinschätzung (Initiierungszeitpunkt II) und die Auswahl zu wiederholender Konzepte durch die Lernenden. Den Lernenden wurde jedoch keine Möglichkeit eingeräumt, die Konzepte tatsächlich erneut zu lernen. In der Folge wurden im Posttest die gleichen Fragen des Pretests gestellt. Abschließend schätzten die Lernenden ihr Abschneiden bei dem Test auf einer dreistufigen Skala (*no credit*, *partial credit*, *full credit*) selbst ein (Initiierungszeitpunkt IV).

Es zeigte sich, dass die relative Akkuratheit der Selbsteinschätzung zum Initiierungszeitpunkt II relativ hoch ist (γ -Korrelation = .78) und der Zusammenhang mit der Wahl der zu wiederholenden Konzepte stark negativ ist (γ -Korrelation = -.80). Eine akkurate Selbsteinschätzung steht demnach in Zusammenhang mit einer guten Selbstregulation, gemessen daran, dass gelernte Inhalte nicht wiederholt werden. Lernende weisen jedoch bei der abschließenden Selbsteinschätzung nach Bearbeitung des Post-Tests (Initiierungszeitpunkt IV) eine hohe Selbstüberschätzung auf, wenn im Pretest inakkurates Vorwissen aktiviert wurde. Entsprechend ist die

Fähigkeit zur Wahl adäquater Items zur Wiederholung weniger ausgeprägt. Van Loon et al. (2013) kommen sowohl für Lernende der dritten und vierten Klasse als auch für jene fünfter und sechster Klassen zu diesem Ergebnis und gelangen somit im außeruniversitären Kontext zu ähnlichen Resultaten wie Dunlosky & Rawson (2012) sowie Thiede et al. (2003).

Die bislang dargestellten Studien hatten zum Untersuchungsgegenstand, wie die Akkuratheit der Selbsteinschätzung mit der Performanz und der Selbstregulation (i. S. der Wahl zu wiederholender Lernaufgaben) zusammenhängt. Darüber hinaus lassen sich einige Studien finden, die diese Zusammenhänge in Bezug auf die Wahl neuer Lernaufgaben betrachten.

Kostons, van Gog, & Paas (2010) untersuchten, inwiefern in einer lernerzentrierten Lernumgebung⁹ Unterschiede zwischen effektiven und ineffektiven Lernenden hinsichtlich der Fähigkeit zur Selbsteinschätzung existieren und welche Zusammenhänge in der Folge mit der Wahl von weiteren Lernaufgaben bestehen. Lernende (N = 32 aus dem Sekundarbereich im Fach Biologie) erhielten im Rahmen einer webbasierten Lernumgebung Aufgaben unterschiedlicher Komplexität und Unterstützungsmerkmale zum gleichen Themengebiet. Im Gegensatz zu den bisher genannten Studien von Dunlosky & Rawson (2012), Thiede et al. (2003) und van Loon et al. (2013) werden hier Aufgaben mit verschiedenen kognitiven Anspruchsniveaus angesprochen.

Nach jeder gelösten Aufgabe schätzten die Lernenden sich in Bezug auf sieben Kriterien (u. a. Richtigkeit der Lösung, Schwierigkeit der Aufgabe und Gesamtbeurteilung des Geleisteten) auf einer neunstufigen Skala selbst ein (Initiierungszeitpunkt II). Nach der Selbsteinschätzung wählten die Lernenden eine neue Lernaufgabe aus und bearbeiteten diese. Um Erkenntnisse bezüglich der von Lernenden herangezogenen Kriterien bei der Selbsteinschätzung und Aufgabenwahl zu gewinnen, wurden Think-aloud-Protokolle erstellt. Diese wurden in Anlehnung an die Taxonomien von Biggs & Collis (1982) kodiert.

Die Ergebnisse zeigen, dass die zuvor im Rahmen eines Pretests als effektive Lerner deklarierten Probanden (n = 17) über eine höhere Akkuratheit in ihrer Selbsteinschätzung verfügen als ineffektive Lerner (n = 15). Gleichzeitig merken Kostons et al. (2010) jedoch an, dass innerhalb der beiden Gruppen (effiziente vs. ineffiziente Lerner) große Varianzen in der Akkuratheit der Selbsteinschätzung bestehen. Lineare Regressionsanalysen wurden durchgeführt, um die Fähigkeit zur Selbsteinschätzung als Prädiktor für die Wahl adäquater Lernaufgaben zu untersuchen. Die unabhängigen Variablen „Richtigkeit der Lösung“, „Zeitaufwand“ und „Gesamtbe-

⁹ Die lernerzentrierte Lernumgebung wurde von Kostons et al. (2010) derart gestaltet, dass Lernende Aufgabenwahl, Zeit zur Aufgabebearbeitung und Lösung der Aufgaben selbst vornahmen.

urteilung des Geleisteten“ sind wichtige Prädiktoren für die Aufgabenwahl der Gruppe der inakkuraten effektiven Lerner. Demnach führen eine geringer eingeschätzte „Richtigkeit der Lösung“ ($\beta = -.78$), ein geringer eingeschätzter „Zeitaufwand“ ($\beta = -.61$) sowie eine höher bewertete „Gesamtbeurteilung des Geleisteten“ ($\beta = .72$) zur Wahl von kognitiv schwierigeren Aufgaben. Für die Gruppe der akkuraten effektiven Lernenden ließ sich ein Modell mit den Parametern „Zeitaufwand“, „Schwierigkeit der Aufgabe“ und „mentale Anstrengung“ erstellen. Ein geringer eingeschätzter „Zeitaufwand“ ($\beta = -.50$), eine geringer beurteilte „Schwierigkeit der Aufgabe“ ($\beta = -.82$) und eine geringer eingeschätzte „mentale Anstrengung“ ($\beta = -.86$) bei der Aufgabenbearbeitung führen dazu, dass Lernende verstärkt schwierigere Aufgaben auswählen.

In einer Folgestudie eruierten Kostons, van Gog, & Paas (2012), inwieweit die Fähigkeit zur akkuraten Selbsteinschätzung sowie der adäquaten Aufgabenwahl trainierbar ist und wie akkurate Selbsteinschätzung, Aufgabenwahl und Performanz zusammenhängen. Hierzu führten sie zwei Experimente durch. Im ersten Experiment konnte die Trainierbarkeit der Fähigkeit zur Selbsteinschätzung nachgewiesen werden. Das Training bestand dabei in der Modellierung des Prozesses der Selbsteinschätzung und der Aufgabenwahl. Den Probanden ($N = 80$ Lernende im Fach Biologie) wurde je nach Treatment-Gruppe¹⁰ das Vorgehen bei der Selbsteinschätzung und/oder der Aufgabenwahl an einem Modell erläutert. Hierbei löste ein Experte beispielhaft eine Aufgabe, schätzte sich selbst ein und wählte eine neue Lernaufgabe aus. Indem der Experte laut dachte, wurden die einzelnen Bearbeitungsschritte expliziert. Im Rahmen eines Pretests bearbeiteten die Lernenden fünf Aufgaben aus dem Bereich Genetik mit unterschiedlichen Schwierigkeitsgraden¹¹. Nach jeder bearbeiteten Aufgabe schätzten sie ihre mentale Anstrengung bei der Aufgabenbearbeitung (neunstufige Skala) sowie ihre Performanz (fünfstufige Skala) selbst ein (Initiierungszeitpunkt II). Anschließend wählten sie eine neue Aufgabe aus einer Aufgabensammlung mit unterschiedlichen Anspruchsniveaus aus. Dabei erhielten sie keine Rückmeldung über Performanz und Akkuratheit der Selbsteinschätzung sowie der Akkuratheit der Aufgabenwahl. Im Anschluss wurde je nach Treatment-Gruppe der Prozess von Selbsteinschätzung und Aufgabenwahl modelliert, bevor die Lernenden im Posttest die gleiche Prozedur wie im Pretest mit lediglich leicht abgeänderten Aufgaben durchliefen.

Es zeigte sich, dass die Modellierung der Selbsteinschätzung in Zusammenhang mit einer signifikant höheren Akkuratheit der Selbsteinschätzung steht, wohingegen sich dieser Effekt bei

¹⁰ Die Probanden wurden dabei randomisiert einer der vier Treatment-Gruppen zugeteilt: i) Modellierung von Selbsteinschätzung und Aufgabenwahl durch Experten, ii) Modellierung der Selbsteinschätzung, iii) Modellierung der Aufgabenwahl und iv) keine beispielhafte Selbsteinschätzung und Aufgabenwahl.

¹¹ Die Einteilung der Aufgaben nach Schwierigkeitsgraden erfolgte dabei analog zu Kostons et al. (2010).

der Aufgabenwahl nicht nachweisen ließ. Die grundlegende Trainierbarkeit der metakognitiven Fähigkeiten sehen Kostons et al. (2012) mit dem ersten Experiment bestätigt.

In einem zweiten Experiment wurden die Zusammenhänge zwischen akkurater Selbsteinschätzung, Selbstregulation und Performanz näher analysiert. Pre- und Posttest sowie die Selbsteinschätzungen zur Performanz und Aufgabenwahl unterschieden sich nicht vom ersten Experiment. Hinzugefügt wurde im zweiten Experiment nach der Betrachtung der modellierten Vorgehensweise zur Selbsteinschätzung und Aufgabenwahl durch einen Experten eine Phase selbstregulierten Lernens. In dieser konnten die Lernenden (N = 90 Studierende) das Prozedere aus Aufgabenbearbeitung, Selbsteinschätzung und Aufgabenwahl an acht selbst ausgewählten Übungsaufgaben trainieren. Kostons et al. (2012) kamen zum Ergebnis, dass die Treatment-Gruppe mit der akkurateren Selbsteinschätzung in einer selbstregulierten Lernphase sowohl eine adäquatere Aufgabenwahl¹² zeigte als auch größere Lernfortschritte erzielte als Probanden mit geringerer Fähigkeit zur Selbsteinschätzung. Eine höhere Akkuratheit der Selbsteinschätzung führte demnach zu einer besseren Selbstregulation und letztlich einer besseren Performanz.

Sitzmann & Johnson (2012) untersuchten für eine Gruppe von Lernenden (N = 293) die Auswirkungen von (in-)akkuraten Selbsteinschätzungen auf die Performanz sowie das Aufsuchen weiterer Lerngelegenheiten. Die Teilnehmenden (zum Großteil Beschäftigte mit unterschiedlichen Bildungshintergründen) partizipierten an einem freiwilligen Online-Excel-Kurs, der aus sechs Modulen bestand. Nach der Bearbeitung eines Moduls erfolgte eine Selbsteinschätzung, indem die Probanden die Zahl der von ihnen richtig gelösten Aufgaben prognostizierten (Initiierungszeitpunkt II). Anschließend bearbeiteten sie einen MC-Test mit Fragen, die auf deklaratives und prozedurales Wissen abzielten. Während der selbstgesteuerten Lernphase wurde die Lernzeit als Indikator für die aufgebrachte Anstrengung gemessen. Es zeigte sich, dass Selbsteinschätzung und Performanz signifikant korrelieren. Die Performanz ist dann am höchsten, wenn diese im vorherigen Modul bereits hoch war und auch eine hohe Selbsteinschätzung vorlag (*uniformly positive ratings*). Überdies ist eine Unterschätzung im Vergleich zu einer Überschätzung nach Sitzmann & Johnson (2012) mit einer höheren Performanz verbunden. Die geringste Performanz ergab sich für Lernende mit geringer Performanz im vorherigen Modul und einer geringen Selbsteinschätzung (*uniformly negative ratings*). Analog dazu gestalten sich die

¹² Hierfür wurden 15 Aufgabentypen in unterschiedliche Komplexitätsstufen und Unterstützungsniveaus klassifiziert. Eine adäquate Aufgabenwahl lag vor, wenn die zur weiteren Bearbeitung empfohlene Niveaustufe der durch den Lernenden ausgewählten Niveaustufe entsprach (Kostons et al., 2012).

Ergebnisse für den Ausstieg aus dem Excel-Kurs. Die höchsten Ausstiegsraten wurden bei signifikanter Überschätzung, gefolgt von *uniformly negative ratings*, Unterschätzung und *uniformly positive ratings* verzeichnet. Ferner ist mit der Anstrengung ein stärkerer Effekt auf die Performanz verbunden, wenn Lernende ihren eigenen Wissensstand für gering erachten. Bei hoher Selbsteinschätzung ist die Performanz unabhängiger von der angegebenen Anstrengung der Lernenden.

In einer zweiten Studie erweiterten Sitzmann & Johnson (2012) ihre Untersuchungen um die Variablen Gewissenhaftigkeit und Einstellung zum Lerngegenstand. In dieser Studie nahmen N = 253 Lernende an einer Excel-Schulung teil, bei der sich die Vorgehensweise analog zu jener der vorangegangenen Untersuchung gestaltete. Es zeigte sich, dass der beobachtete Zusammenhang zwischen Selbsteinschätzung und Performanz mit der Ausprägung der Gewissenhaftigkeit der Probanden variiert. Eine Fehleinschätzung der eigenen Performanz führt eher zu einem Ausscheiden aus der Excel-Schulung, wenn die Gewissenhaftigkeit der Probanden gering ausgeprägt ist. Dagegen verschlechtert sich die Performanz bei einer Fehleinschätzung besonders für Probanden mit hoher Gewissenhaftigkeit. Die Zusammenhänge werden dabei vom *commitment* der Lernenden mediiert, wobei das *commitment* einen größeren negativen Effekt auf das Ausscheiden aus der Schulung hatte, wenn die Selbsteinschätzung der Probanden niedrig ausgeprägt war.

Panadero & Romero (2014) betrachteten die Effekte des Einsatzes von Rubrics zur Selbsteinschätzung i. Vgl. zur einfachen Aufforderung zur Selbsteinschätzung der Performanz auf die Selbstregulation, die Performanz, die Akkuratheit der Selbsteinschätzung und den aufgabenbezogenen Stress. Ziel war es, die dem Einsatz von Rubrics zugeschriebenen positiven Wirkungen empirisch zu überprüfen. Die Probanden (N = 218 angehende Lehrpersonen) wurden aufgefordert, eine graphische Zusammenfassung zu einem Text zu erstellen. Die Qualität der Abbildung sollten diese schließlich selbst einschätzen (Initiierungszeitpunkt II). Dabei erhielt die Treatment-Gruppe (n = 111) bereits vor Beginn der Erarbeitungshase von Experten entwickelte Rubrics, die sämtliche für die Beurteilung relevanten Kriterien enthielten. Die Kontrollgruppe (n = 107) erhielt lediglich eine Aufforderung zur Selbsteinschätzung des Geleisteten. Die Ausprägung der Selbstregulation wurde durch Einsatz des *Emotion and Motivation Self-Regulation Questionnaire* (Alonso Tapia, Panadero, & Ruiz, 2014) sowie der Möglichkeit der freien Beantwortung durch die Lernenden auf die Fragestellung, welche Strategien zur Aufgabenbewältigung eingesetzt wurden, erfasst. Panadero & Romero (2014) gelangen zum Ergebnis, dass die Lernenden in der Treatment-Gruppe über eine akkuratere Selbsteinschätzung und eine bessere Performanz verfügen. Die Erwartung, dass eine bessere Fähigkeit zu selbstreguliertem Lernen

den Zusammenhang zwischen der Güte der Selbsteinschätzung und der Performanz erklären könnte, wurde nicht bestätigt. Die Treatment-Gruppe wies für beide Unterskalen (*Learning Self-Regulation* und *Performance/Avoidance Self-Regulation*) höhere Werte auf als die Kontrollgruppe. Zu erwarten war an dieser Stelle, dass Lernende mit besserer Selbsteinschätzung tendenziell höhere Werte in der *Learning-Self-Regulation*-Skala und niedrigere Werte bei der *Performance/Avoidance Self-Regulation* aufweisen. Dieses Ergebnis ist nicht kongruent zur Argumentationskette, dass eine akkuratere Selbsteinschätzung des Gelernten über eine bessere Selbstregulation im Lernprozess zu einer verbesserten Performanz im Rahmen eines abschließenden Tests führt.

3.2.2 Effekte (in-)akkuratere Selbsteinschätzungen auf die Performanz

In diesem Kapitel werden Studien angeführt, die auf der einen Seite die Akkuratheit der Selbsteinschätzung erfassen und diese auf der anderen Seite in Zusammenhang mit der Performanz bringen. Die Studien unterscheiden sich von jenen in Kapitel 3.2.1 dahingehend, dass die potenziellen Zusammenhänge mit der Selbstregulation (i. S. der Wahrnehmung von weiteren Lerngelegenheiten) unberücksichtigt bleiben. Tabelle 2 gibt einen Überblick über die hier skizzierten Studien.

Studie	Selbsteinschätzung	Initiierungszeitpunkt	Messung der Performanz
Nietfeld, Cao, & Osborne (2006)	Judgements of Learning Confidence Judgements (wöchentlich)	II & III	MC-Aufgaben & offene Aufgaben (Schema-Repräsentation)
Mengelkamp & Bannert (2010)	Confidence Judgements	III & IV	Wissenstests mit verschiedenen Aufgabenniveaus
Snyder, Nietfeld, & Linnenbrink-Garcia (2011)	Judgements of Learning Confidence Judgements	II & III	4 Prüfungen in Biologie (MC-Aufgaben & andere Formate)
Maki, Jonas, & Kallod (1994)	Judgements of Learning Confidence Judgements	II & III	Verschiedene Tests zum Textverständnis
Coutts & Rogers (1999)	Confidence Judgements	III	Verschiedene Testinstrumente
Olina & Sullivan (2004)	Confidence Judgements	III	Projektbericht Test
Hewitt (2002)	Judgements of Learning	II	Lernen am Modell Woodwind Brass Solo Evaluation Form
Glenberg & Epstein (1985)	Confidence Judgements	II & III	Prüfung von Aussagen auf ihren Wahrheitsgehalt
Begg, Martin, & Needham (1992)	Judgements of Learning	II	Testung des Erinnerungsvermögens beim Lernen von Wortpaaren
Hartwig, Was, Isaacson, & Dunlosky (2012)	Judgements of Learning	II	Test zu Vokabeln (MC-Aufgaben & komplexere Formate)

Tabelle 2: Überblick über Studien der zweiten Kategorie

Nietfeld, Cao, & Osborne (2006) untersuchten über den Einsatz von Aufgaben mit unterschiedlichen Anspruchsniveaus, inwiefern Übungen und Feedback zur Selbsteinschätzung einen Effekt auf die Akkuratheit der Selbsteinschätzung, die Performanz und die Selbstwirksamkeit der Lernenden besitzen. Sie betrachten dabei explizit, inwiefern eine gut angepasste Kalibrierung (d. h. eine gute Passung von Wahrnehmung der eigenen Performanz und tatsächlich realisierter Performanz) Änderungen in der Performanz bewirkt. Die Treatment-Gruppe (n = 45 Studierende der Pädagogischen Psychologie) bearbeitete am Ende einer jeden Seminarsitzung kurze

Übungen zum Monitoring¹³. Diese beinhalteten Selbsteinschätzungen auf einer kontinuierlichen Skala von 0 bis 100 % zur Überzeugung (Initiierungszeitpunkt II), das zuvor Gelernte wiedergeben zu können, sowie die Reflexion über Schwierigkeiten im Umgang mit den Lerninhalten. Ferner wurden von den Lernenden Wiederholungsfragen bearbeitet, deren Antworten ebenfalls auf Richtigkeit selbst eingeschätzt wurden (Initiierungszeitpunkt III). Die Lernenden erhielten dabei Feedback zur Akkuratheit ihrer Selbsteinschätzung vom Dozenten. Die Kontrollgruppe (n = 39 Studierende der Pädagogischen Psychologie) wurde ebenfalls angehalten, Selbsteinschätzungen vorzunehmen, bearbeitete jedoch keine Übungen und erhielt somit auch kein Feedback. Ansonsten unterlagen beide Gruppen den gleichen Bedingungen (gleicher Dozent, gleiche Seminarinhalte). Im Laufe des Semesters wurden neben Pre- und Post-Test vier weitere Tests eingesetzt, um die Entwicklung der Akkuratheit der Selbsteinschätzung und der Performanz zu analysieren. Ferner wurde zur Performanzmessung auf ein Projekt der Schema-Repräsentation zurückgegriffen. Hiermit wurde beabsichtigt, ein mentales Modell der Studierenden zu effektivem Unterricht zu externalisieren.

Die Treatment-Gruppe weist im Zeitverlauf eine bessere Kalibrierung auf als die Kontrollgruppe. Da diese im Pretest noch für beide Gruppen ähnlich war, wird der Zuwachs der akkuraten Selbsteinschätzung der Intervention zugeschrieben. Analog dazu zeigen sich für die Treatment-Gruppe signifikant bessere Werte in der Performanz, gemessen in den vier Tests im Semesterverlauf. Ein Pfadmodell wurde entwickelt, um die Bedeutung der Kalibrierung und der Veränderung der Kalibrierung im Zeitverlauf für die Performanz festzustellen. Ergebnisse multipler Regressionsanalysen zeigen, dass der Anstieg in der Kalibrierung einen substantiellen Effekt auf die Performanz im finalen Test hat ($\beta = .47^{***}$). Nietfeld et al. (2006) kommen zum Schluss, dass die Kalibrierung ein wichtiger Indikator für die Performanz ist. Darüber hinaus zeigen die Ergebnisse hinsichtlich der Schema-Repräsentation, dass die Beobachtungen nicht nur für MC-Aufgaben, sondern auch für offene Aufgabenformate gemacht werden können, für die vielfältige Lösungsmöglichkeiten kennzeichnend sind. Demnach führt ein akkurateres Monitoring zu einem besseren Abschneiden in offenen Aufgabensituationen.

Mengelkamp & Bannert (2010) untersuchten vorrangig die Stabilität und die Generalisierbarkeit von verschiedenen Instrumenten zur Messung der Akkuratheit der Selbsteinschätzung¹⁴

¹³ Nietfeld et al. (2006) verstehen unter Monitoring vordergründig das Bewusstsein des Lernenden über Verständnis und Performanz zu einer Aufgabe. Dies sei zentral für die Steuerung des Lernprozesses.

¹⁴ Als Maß für die Akkuratheit der Selbsteinschätzung führen Mengelkamp & Bannert (2010) das Konzept des Bias an. Hierunter wird die Differenz zwischen durchschnittlicher Selbsteinschätzung und durchschnittlicher Performanz verstanden. Das absolute Bias ignoriert dabei die Richtung der Fehleinschätzung, indem die Differenzen bspw. quadriert werden.

(absolute vs. relative Akkuratheit) und bilden neben der Studie von Nietfeld et al. (2006) eine der wenigen Längsschnittuntersuchungen in diesem Kontext ab. Dabei betrachteten sie auch die prädiktive Validität der Instrumente bezogen auf die Performanz. Studierende der Psychologie und Erziehungswissenschaften ($N = 113$) bearbeiteten im Rahmen einer virtuellen Lernumgebung (Hypermedia) verschiedene Informationen zum Themengebiet des „Operanten Konditionieren“. Hierbei wurden drei Wissenstests eingesetzt, die Aufgaben mit unterschiedlichen kognitiven Anspruchsniveaus enthielten (Wissen, Verstehen und Transfer). Das Abschneiden bezüglich dieser Aufgaben schätzten die Studierenden durch Angabe der prozentualen Wahrscheinlichkeit der Richtigkeit der Antwort ein (Initiierungszeitpunkt III & IV). Es zeigte sich, dass die Studierenden in der Lage waren, ihre Leistungen genau einzuschätzen. Diese Fähigkeit war im Zeitverlauf bezogen auf die absolute Akkuratheit stabil. Signifikante Korrelationen weisen überdies darauf hin, dass die Genauigkeit der Selbsteinschätzung über verschiedene Tests hinweg generalisierbar ist. Die prädiktive Validität ist für Instrumente der relativen Akkuratheit gegeben. Das absolute Bias war ein signifikanter Prädiktor für die Performanz der Lernenden ($r = -.42^{***}$). Für das relative Bias ließ sich dieser Zusammenhang jedoch nicht feststellen ($r = -.18$, n. s.).

Snyder, Nietfeld, & Linnenbrink-Garcia (2011) betrachteten die Unterschiede in der Metakognition (i. S. von spezifischen und pauschalen Confidence Judgements) zwischen normalbegabten und hochbegabten Lernenden. Es wurde vermutet, dass hochbegabte Lernende über bessere Fähigkeiten zur Kalibrierung verfügen, genauere Selbsteinschätzungen entwickeln und somit bessere Lernergebnisse in Leistungstests erreichen. Insgesamt fünf homogene Klassen ($N = 67$ Schülerinnen und Schüler) hochbegabter und normalbegabter Schülerinnen und Schüler bearbeiteten im Verlauf eines Schulhalbjahres vier Prüfungen, die verschiedene Aufgabenformate beinhalteten (z. B. MC-Aufgaben oder Ergänzungsaufgaben). Bei der Bearbeitung der Prüfungen wurden diverse Instrumente zur Erfassung der Akkuratheit der Selbsteinschätzung eingesetzt. Vor und nach den Prüfungen gaben die Lernenden an, wie viele Aufgaben sie vermutlich richtig lösen werden bzw. richtig gelöst haben (globale Einschätzungen, Initiierungszeitpunkt II & III). Darüber hinaus schätzten sie auf einer kontinuierlichen Skala mit einem Wertebereich von 0 bis 100 für jede Aufgabe die Wahrscheinlichkeit ein, mit der sie glauben, die Aufgabe korrekt gelöst zu haben (lokale Einschätzungen, Initiierungszeitpunkt III). Ferner wurde mit dem *Metacognitive Awareness Inventory* (MAI) von Schraw & Dennison (1994) ein Instrument (52 Items) zur Messung des Wissens über Kognitionen (metakognitive Strategien, Fähigkeit zum Einsatz der Strategien, etc.) und der kognitiven Regulation (Planung, Monitoring und Evaluation) eingesetzt. Hinsichtlich der Ausprägungen des MAI konnten Snyder et al. (2011) keine

signifikanten Unterschiede zwischen normal- und hochbegabten Lernenden feststellen. Erwartungsgemäß übertrafen die hochbegabten Lernenden die normalbegabten in jeder der vier Prüfungen in der Performanz. Gleichzeitig wiesen sie ein akkurateres Monitoring auf. Allerdings konnten für die Voraussage der Anzahl der richtig gelösten Items keine Unterschiede zwischen den Gruppen gefunden werden. Auf Basis multipler Regressionsanalysen untersuchten Snyder et al. (2011) die Zusammenhänge zwischen lokalen Selbsteinschätzungen und der Performanz in Abhängigkeit der Begabung der Lernenden. Es zeigte sich, dass die besseren Performanzwerte der hochbegabten Lernenden sich teilweise durch die bessere Akkuratheit der Kalibrierung erklären ließen.

Maki, Jonas, & Kallod (1994) untersuchten die Zusammenhänge zwischen verschiedenen Instrumenten zur Messung des Leseverständnisses und der Fähigkeit, die Performanz in einem Test vorherzusagen (Initiierungszeitpunkt II) bzw. nach einem Test einzuschätzen (Initiierungszeitpunkt III). Hierbei lassen sich Rückschlüsse über den Zusammenhang von Akkuratheit der Selbsteinschätzung und Performanz ziehen. 75 Studierende der Psychologie bearbeiteten im Rahmen dieser Studie Texte, die mit dem Einsatz der *Multimedia Comprehension Battery* (MMCB)¹⁵ und des *Nelson-Denny-Reading Test*¹⁶ vorgesehen sind. Zwei weitere Texte wurden eingesetzt, um das Textverständnis zu prüfen. Die Lernenden wurden angehalten, die Vorhersage der Performanz nach Bearbeitung der Texte sowie nach Bearbeitung des Tests auf einer sechsstufigen Skala vorzunehmen. Es zeigte sich, dass die Vorhersage der Performanz vor Bearbeitung des Tests inakkurat ausfiel (γ -Korrelation von .114), während die Vorhersage nach der Bearbeitung des Tests akkurater gewesen ist (γ -Korrelation von .551). Darüber hinaus ergaben sich fast durchgängig signifikante Korrelationen zwischen der Selbsteinschätzung nach der Testung (Initiierungszeitpunkt III) und der Performanz, was zum Schluss führt, dass Probanden mit einer besseren Akkuratheit der Selbsteinschätzung auch eine bessere Performanz im abschließenden Test zeigen.

Coutts & Rogers (1999) analysierten in ihrer Studie an einer Stichprobe von 420 Studierenden der Medizin, welche Prädiktoren für eine akkurate Selbsteinschätzung bestehen, worin sich Lernende, die zur Überschätzung neigen, von Lernenden, die sich tendenziell unterschätzen, unterscheiden und welche Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der Performanz bestehen. Hierzu bearbeiteten die Lernenden diverse medizinische Fälle

¹⁵ Die MMCB unterscheidet das Verständnis von schriftlichen, auditiven und piktoralen Inhalten.

¹⁶ Der Nelson-Denny-Reading Test ist ein Leseverständnistest, bestehend aus acht expositorischen Texten, sich darauf beziehende MC-Fragen sowie der Messung der Leserate (Anzahl gelesener Worte innerhalb einer Minute).

anhand eines Testpatienten, der charakteristische Symptome simulierte, und schätzten ihre Performanz auf einer fünfstufigen Skala selbst ein (Initiierungszeitpunkt III). Die Ergebnisse zeigen, dass ein Großteil der Lernenden eine akkurate Selbsteinschätzung aufwies (86.8 %, bei einem Cut-Off von 20 %¹⁷). Nur wenige Lernende überschätzten (5.3 %) bzw. unterschätzten (7.9 %) ihre Performanz. Es zeigte sich, dass Lernende, die eine akkurate Selbsteinschätzung bzw. Unterschätzung aufwiesen, höhere Werte bei der Performanz erzielten als Lernende, die sich überschätzten.

Olina & Sullivan (2004) untersuchten, welchen Effekt die Kombination von Selbsteinschätzung durch die Lernenden und Fremdeinschätzung durch die Lehrperson auf die Performanz besitzt. Ferner wurde betrachtet, welchen Einfluss die Assessment-Verfahren auf die Einstellungen der Lernenden haben. 341 Lernende zehnter und elfter Klassen durchliefen eine zwölfstündige Lerneinheit zur Einführung in die empirische Forschung, in der sie einen Projektbericht erstellen und abschließend einen Performanztest absolvieren mussten. Die Versuchspersonen wurden auf vier Treatment-Gruppen (keine Bewertung im Lernprozess, zwei Selbsteinschätzungen im Laufe der Lerneinheit, zwei Lehrereinschätzungen zum Gelernten im Laufe der Lerneinheit sowie zwei Selbst- und Lehrereinschätzungen im Lauf der Lerneinheit) aufgeteilt. Die acht Lehrpersonen erhielten ausgearbeitete Instruktionsmaterialien zur Einführung in die empirische Forschung. Die Selbst- und Lehrereinschätzungen (Initiierungszeitpunkt III) erfolgten anhand einer projektbezogenen Skala von 15 Items, die auf einer dreistufigen Skala eingeschätzt werden mussten. Zur Bestimmung der Einstellung der Lernenden und Lehrenden wurden Skalen von sechs bzw. elf Items eingesetzt. Olina & Sullivan (2004) kommen zum Ergebnis, dass Lernende, die sich sowohl selbst einschätzen mussten als auch eine Einschätzung der Lehrperson erhielten, die höchsten Performanzwerte für die Erstellung des Projektberichtes erreichten. Für den Wissenstest im Rahmen des Posttests konnten jedoch keine signifikanten Unterschiede zwischen den Treatments ermittelt werden. Lernende, die eine Selbsteinschätzung vornehmen mussten, wiesen höhere Einstellungswerte hinsichtlich der Zufriedenheit mit dem Lernarrangement, der Selbstüberzeugung und des wahrgenommenen Lernzuwachses auf. Zuletzt betrachteten Olina & Sullivan (2004) die Akkuratheit der Selbsteinschätzung. Lernende neigen demnach zur Überschätzung der eigenen Performanz.

¹⁷ Die Probanden wurden in Abhängigkeit von der Akkuratheit der Selbsteinschätzung in drei Gruppen eingeteilt. Lagen die Selbsteinschätzungen mehr als 20 % über der tatsächlichen Performanz, wurden Probanden der Gruppe der Überschätzer zugewiesen. In einem Intervall von +/- 20 % von der tatsächlichen Performanz wurde von einer akkuraten Selbsteinschätzung ausgegangen, im Falle einer Abweichung von über 20 % unterhalb der tatsächlichen Performanz von Unterschätzung (Coutts & Rogers, 1999, S. 129).

Hewitt (2002) untersuchte die Fähigkeit zur akkuraten Selbsteinschätzung von Lernenden beim Erlernen von Musikinstrumenten. Ziel war es, die Entwicklung der Selbsteinschätzung im Zeitverlauf zu eruieren, die Akkuratheit der Selbsteinschätzung in Abhängigkeit unterschiedlicher Treatments zu bestimmen sowie Zusammenhänge zwischen akkurater Selbsteinschätzung und der Performanz zu analysieren. Vor dem Hintergrund der Modellierung selbstregulierten Lernens nach Zimmerman (2000) wurde erwartet, dass eine akkurate Selbsteinschätzung positiv mit der Performanz zusammenhängt. Lernende siebter bis neunter Klassen ($N = 41$) wurden im regulären sechswöchigen Unterricht, in dem das Spielen eines Instrumentes gelernt werden sollte, angewiesen, ausgewählte Musikstücke zu erlernen. Dabei wurden sie in zwei Treatment-Gruppen eingeteilt. Die erste Gruppe erhielt die Möglichkeit, Audioaufnahmen über das akkurate Spielen der Stücke zu hören, während die zweite Gruppe diese Aufnahmen nicht hörte. Die Selbsteinschätzung (Initiierungszeitpunkt II) sowie die Erfassung der Performanz erfolgte mit Einsatz der *Woodwind Brass Solo Evaluation Form*, mit dem Kriterien festgelegt sind, die als Standard für die Performanzmessung herangezogen werden können. Es zeigte sich, dass die Selbsteinschätzung im Zeitverlauf stieg, das Treatment jedoch keinen Einfluss auf die Akkuratheit dieser hatte. Hewitt (2002) ermittelte stattdessen, dass Lernende dazu neigen, ihre Performanz zu überschätzen. Letztlich konnten nur für zwei von sieben Kriterien signifikante Zusammenhänge zwischen der Selbsteinschätzung der Lernenden und der Experteneinschätzung festgestellt werden. Es bestehen außerdem keine Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der Performanz.

Im Rahmen von drei Experimenten untersuchten Glenberg & Epstein (1985) an mehreren, aus Studierenden der Psychologie bestehenden Stichproben, inwieweit die Lernenden in der Lage sind, Gelerntes und nicht Gelerntes zu erkennen und ihre Selbsteinschätzung der tatsächlichen Performanz anzupassen (Kalibrierung). In den ersten beiden Experimenten werden Zusammenhänge zwischen Selbsteinschätzung (Initiierungszeitpunkt II) und der Performanz geprüft. Im ersten Experiment wurden die Lernenden angehalten, Kurztexte zu lernen und anschließend die erwartete Performanz in einem Test einzuschätzen. Unterschieden wurden zwei Treatment-Gruppen hinsichtlich des Zeitpunktes der Selbsteinschätzung (*immediate vs. delayed Confidence Judgements*). Die erste Treatment-Gruppe schätzte sich unmittelbar nach Bearbeitung eines Textes selbst ein, die zweite erst, nachdem alle Texte bearbeitet wurden. Die Vorgehensweise ist somit mit jener von Thiede et al. (2003) vergleichbar. Die Selbsteinschätzung wurde anhand einer sechsstufigen Skala vorgenommen. Es zeigte sich, dass Lernende nicht in der Lage

waren, zu erkennen, welche Aufgaben verstanden wurden und welche nicht¹⁸. Hierin unterschieden sich die beiden Treatment-Gruppen nicht signifikant. Im zweiten Experiment wurde getestet, inwiefern die Ergebnisse aus dem ersten Experiment von der Vertrautheit im Umgang mit den Testaufgaben abhingen. Um dem entgegenzuwirken, wurden die Lernenden mit Übungen bzw. durch bloßes Lesen der Beispielaufgaben mit dem Testungsverfahren trainiert. Nichtsdestotrotz zeigten sich ähnliche Ergebnisse wie im ersten Experiment; die Lernenden wiesen eine geringe Fähigkeit zu akkuraten Selbsteinschätzung auf. Im dritten Experiment wurde das Setting komplexer gestaltet. Die Lernenden mussten zunächst erneut eine Selbsteinschätzung zum Verständnis des gelesenen Textes abgeben (Anfangskalibrierung; Initiierungszeitpunkt II). Anschließend erfolgte die Testung, gefolgt von einer Selbsteinschätzung der Performanz (Kalibrierung zur Performanz, Initiierungszeitpunkt III). Daran anknüpfend wurden die Selbsteinschätzungen des ersten Schrittes überdacht und eine weitere Testaufgabe gestellt (Re-Kalibrierung). Die zu Beginn erwartungsgemäß gering ausgeprägten Fähigkeiten zur Selbsteinschätzung verbesserten sich im Zeitverlauf. Allerdings lassen sich keine Steigerungen in der Performanz der Lernenden beobachten (durchschnittlich 78 % bzw. 79 % korrekte Antworten). Höhere Akkuratheitswerte sind in diesen Studien also nicht mit höheren Performanzwerten verbunden.

Auch Begg, Martin, & Needham (1992) zweifeln wie Glenberg & Epstein (1985) an der Nützlichkeit akkurater Selbsteinschätzungen bezogen auf das Lernergebnis. Sie untersuchten im Rahmen zweier Studien die Effekte von (in-)akkuratem Monitoring auf das Lernen von Wortpaaren. Hierzu wurden Studierenden der Psychologie (N = 43 bzw. N = 86) 96 Wortpaare für je vier Sekunden zum Lernen präsentiert. Anschließend sollten die Lernenden auf einer fünfstufigen Skala die Wahrscheinlichkeit einschätzen, mit der sie die Wortpaare unter verschiedenen Bedingungen¹⁹ in einem finalen Test wiedergeben können (Initiierungszeitpunkt II). Sie kommen zum Ergebnis, dass die Selbsteinschätzung nicht hilfreich für die Lernenden im Sinne einer effizienteren Allokation der Lerngelegenheiten gewesen ist und auch keinen Einfluss auf die Performanz im abschließenden Test hatte. Diese Befunde konnten Begg et al. (1992) im Rahmen einer zweiten Studie bestätigen.

¹⁸ Hierzu wurden drei Maße berechnet: *Point biserial correlations*, *Confidence Judgement Accuracy Quotient* & *Calibration Curves*.

¹⁹ Die Bedingungen waren *cue review* (Hilfestellung durch Angabe eines Wortes des Wortpaares) und *pair review* (Angabe des Wortpaares).

Hartwig, Was, Isaacson, & Dunlosky (2012) untersuchten im Unterschied zu den bisher referierten Studien, inwiefern akkurates Monitoring und Performanz unabhängig von der Lernaufgabe zusammenhängen. Im Kern wird analysiert, ob eventuell eine generelle Fähigkeit zum akkuraten Monitoring besteht, mit der die Performanz domänenübergreifend vorhergesagt werden kann. Studierende eines Einführungskurses in die Pädagogische Psychologie ($N = 379$) wurden zur Bestimmung der Akkuratheit des Monitorings zunächst gefragt, inwiefern sie die Bedeutung von 50 Vokabeln²⁰ kennen (Initiierungszeitpunkt II). Anschließend wurden die Probanden zu den 50 Vokabeln getestet, wobei sie fünf Antwortalternativen zur Bedeutung der jeweiligen Vokabel erhielten. Die Akkuratheit der Selbsteinschätzung bestimmten Hartwig et al. (2012) über die Berechnung von γ -Korrelationswerten. Zum Semesterende wurde eine abschließende Prüfung, bestehend aus 100 Items²¹ zu den Seminarinhalten aus dem ganzen Semester, durchgeführt. Hartwig et al. (2012) analysierten schließlich die Zusammenhänge zwischen Akkuratheit des Monitorings und der Performanz. Sie ermittelten eine schwache, jedoch signifikante Korrelation von $r = .26$ zwischen Monitoring und Performanz beim Abschlusstest. Die Lernenden mit höheren Werten bei der Akkuratheit der Selbsteinschätzung erzielten auch bessere Werte im abschließenden Test auf einem anderen Themengebiet. Die Akkuratheit der Selbsteinschätzung erklärt jedoch lediglich 6.8 % der Varianz der Performanz. Hartwig et al. (2012) weisen auch einen positiven Zusammenhang zwischen der verbalen Fähigkeit der Lernenden, gemessen am Abschneiden des Vokabeltests, und dem Testergebnis an Semesterende nach ($r = .39$). Allerdings blieb auch bei Kontrolle der verbalen Fähigkeiten der Zusammenhang zwischen akkuratem Monitoring und Performanz in einem anderen Themenbereich bestehen.

²⁰ Von den 50 Vokabeln wiesen 33 Vokabeln einen Themenbezug zu den Seminarinhalten auf, 17 Vokabeln waren gänzlich unabhängig von den Seminarinhalten.

²¹ Diese umfassten Aufgaben unterschiedlicher Anspruchsniveaus. 40 Aufgaben zielten auf Reproduktion, 40 Aufgaben auf Anwendung und 20 Aufgaben auf Integration des Wissens ab.

3.2.3 Effekte (in-)akkuratere Selbsteinschätzungen auf weitere lernbedeutsame Konstrukte

Neben den Studien, die sich mit den Zusammenhängen zwischen der Akkuratheit der Selbsteinschätzung und der Performanz auseinandersetzen, bestehen solche, die anstelle der Performanz weitere lernbedeutsame Konstrukte untersuchen. Tabelle 3 gibt einen Überblick hierüber.

Studie	Selbsteinschätzung	Initiierungszeitpunkt	Messung der Performanz	Lernbedeutsame Konstrukte
Narciss, Koerndle, & Dresel (2011)	Ease of Learning Judgements ~ Judgements of Learning	I & II	4 Tests à 15 Rechenaufgaben	Zufriedenheit mit der Performanz
Nietfeld, Cao, & Osborne (2006)	Judgments of Learning Confidence Judgements	II & III	Test	Selbstwirksamkeitserwartung
Marshall & Brown (2006)	~ Ease of Learning Judgements Confidence Judgements	I & III	Assoziations-test	Emotionales Wohlbefinden

Tabelle 3: Überblick über Studien der dritten Kategorie

Narciss, Koerndle, & Dresel (2011) untersuchten die Auswirkungen der Akkuratheit der Selbsteinschätzungen auf die Zufriedenheit der Lernenden mit ihrer Performanz. Lernende fünfter Klassen ($N = 67$) bearbeiteten wöchentlich Mathematikaufgaben anhand einer eigens entwickelten Mathematiksoftware (MatheWarp) und schätzten sowohl ihre Performanz als auch ihre Zufriedenheit auf einer bipolaren zehnstufigen Skala ein (Initiierungszeitpunkt I & II). Als Interventionsmaßnahme erhielten Probanden der ersten Treatment-Gruppe Feedback zur aktuellen Performanz. Das Feedback erfolgte im Rahmen einer summativen Beurteilung und umfasste die Zahl der richtig gelösten Aufgaben. Die zweite Treatment-Gruppe bekam überdies eine Rückmeldung bzgl. der Attribuierung/Anstrengung bei der Aufgabenbearbeitung (z. B. *This time your effort investment was outstanding. Your success indicates, it was worth the effort* (Narciss et al., 2011, S. 233)). Im Pre- und Posttest wurden je 4 Tests à 15 Rechenaufgaben eingesetzt. Es zeigte sich, dass die Treatment-Gruppen eine bessere Performanz zeigten als die Kontrollgruppe, die Fähigkeit zur Selbsteinschätzung jedoch stabil blieb. Lernende, die ihre Leistung überschätzten, waren gleichzeitig zufriedener. Narciss et al. (2011) sehen in ihrer Studie die soziokognitive Perspektive bestätigt, nach der die Vorteile einer Überschätzung überwiegen.

Nietfeld, Cao, & Osborne (2006) betrachteten neben den Zusammenhängen von Akkuratheit der Selbsteinschätzung und Ausprägung der Performanz die Auswirkungen von Selbsteinschätzungen auf die Selbstwirksamkeit der Lernenden²². Es zeigte sich, dass die durchschnittliche Fähigkeit zur Kalibrierung, also der Passung zwischen der eigenen Wahrnehmung des Gelernten und des tatsächlich Gelernten, einen höchst signifikanten Effekt ($\beta = .44^{***}$) auf die Selbstwirksamkeit der Lernenden in einem abschließenden Test aufweist.

Marshall & Brown (2006) untersuchten in ihren Studien den Zusammenhang zwischen den Erwartungen der Lernenden hinsichtlich ihrer Performanz im Rahmen eines Tests und ihrem emotionalen Wohlbefinden. Ziel war es, die verbreitete Annahme zu prüfen, dass Lernende mit geringeren Erwartungen an das erreichte Lernergebnis positivere Emotionen bzgl. des Lernens aufweisen als Lernende mit hohen Erwartungen. Hierzu wurden $N = 81$ Studierende gebeten, einen computergestützten Assoziationstest (sinngemäße Ergänzung von Wortreihen) zu bearbeiten. Zuvor beantworteten sie einen aus vier Subskalen²³ bestehenden Fragebogen zu ihren Emotionen. Nachdem den Lernenden Beispielaufgaben präsentiert worden sind, gaben sie an, wie viele Aufgaben dieser Art sie richtig zu lösen erwarten (Initiierungszeitpunkt I). Anschließend wurden die Aufgaben bearbeitet. Den Schlusspunkt der Untersuchung bildete die Selbsteinschätzung der Performanz auf einer neunstufigen Skala (Initiierungszeitpunkt III) verbunden mit der erneuten Angabe über gefühlte Emotionen. Hierbei wurde das gleiche Instrument verwendet wie zu Beginn der Testung.

Marshall & Brown (2006) kommen zum Ergebnis, dass Lernende mit hohen Erwartungen an die eigene Performanz auch positive Emotionen erleben. Die Erwartungen sind ein signifikanter Prädiktor für die Ausprägung des eigenen Selbstwertgefühls (*Feelings of Self-Worth*). Es lassen sich nur wenige Hinweise finden, dass eine niedrige Erwartung in der Folge positivere Emotionen auslöst. So bilden geringe Erwartungen an die eigene Performanz lediglich einen signifikanten Prädiktor für das Erleben von Überraschung. Für die anderen Subskalen besteht dieser Zusammenhang nicht. Marshall & Brown (2006) erklären das Ergebnis damit, dass grundlegende positive emotionale Tendenzen die beobachteten Zusammenhänge bedingen. Darüber hinaus weisen Lernende mit hohen Erwartungen vor der Testung, aber niedrigen Performanzwerten, höhere Werte bei der Selbsteinschätzung nach der Testung auf. Lernende mit geringer Performanz und niedrigen Erwartungen schätzten sich nach der Testung schlechter ein als Lernende mit positiven Erwartungen.

²² Für Angaben zum Studiendesign s. Kapitel 3.2.2.

²³ Die Subskalen sind *Feelings of Self-worth*, *Happiness-Sadness*, *Agitation-Relaxation* und *Surprise* (Marshall & Brown, 2006, S. 46).

In einer weiteren Studie ergänzten Marshall & Brown (2006) ihre Untersuchung um den Aspekt der Attribuierung. Geprüft wurden die Zusammenhänge zwischen den Erwartungen der Studierenden, der Performanz und der Attribuierung. Das Design der Studie glich der ersten beschriebenen Teilstudie von Marshall & Brown (2006). Es bestätigte sich das Ergebnis der ersten Studie, dass hohe Erwartungen mit positiveren Emotionen einhergehen, unabhängig davon, ob diese berechtigt oder unberechtigt gewesen sind.

3.3 Diskussion der Forschungsbefunde

3.3.1 Zusammenfassung der Ergebnisse

Mit Kapitel 3.2 wurde das Ziel verfolgt, die aktuelle empirische Befundlage zu den Effekten des Einsatzes von Self-Assessments zu eruieren. In bisherigen Reviews zum Self-Assessment wurde der Fokus auf die Akkuratheit der Selbsteinschätzung, die Bedingungsfaktoren akkurater Selbsteinschätzungen sowie auf die Effekte, die mit dem Einsatz von Self-Assessments verbunden sind, gesetzt (Brown & Harris, 2013; Dochy et al., 1999; Falchikov & Boud, 1989). Nicht berücksichtigt wurden dabei die Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung, dem weiteren selbstregulierten Lernen sowie der Performanz.

Aus diesem Grund wurden in einem ersten Schritt Studien identifiziert, in denen die Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung, der Selbstregulation des Lernens und der Performanz untersucht worden sind. Die Ergebnisse der recherchierten Studien erweisen sich trotz unterschiedlicher methodischer Vorgehensweisen und differierender Bildungshintergründe der Probanden (Primarschülerinnen und -schüler bis zu Universitätsstudierenden) als recht konsistent. Die angenommenen positiven Zusammenhänge zwischen einer akkuraten Selbsteinschätzung, selbstreguliertem Lernen und besserer Performanz bestätigen sich in den meisten Studien. Sowohl im universitären (z. B. Dunlosky & Rawson, 2012; Panadero & Romero, 2014; Thiede et al., 2003) als auch im schulischen Kontext (z. B. Kostons et al., 2010, 2012; van Loon et al., 2013) zeigte sich, dass die Lernenden mit einer inakkuraten Selbsteinschätzung ihren Lernprozess weniger effizient gestalten als Lernende mit einer akkuraten Selbsteinschätzung, indem sie nach einer Selbsteinschätzung eher Lernaufgaben zum weiteren Lernen auswählen, die ihrem Wissensstand nicht entsprechen. Dies schlägt sich auch in den Ergebnissen abschließender Performanztests nieder, in denen Lernende mit besseren Werten bzgl. der Akkuratheit bessere Ergebnisse erzielen. Die Studienbefunde sind dabei unabhängig vom Initiierungszeitpunkt der Selbsteinschätzung. Sowohl für Selbsteinschätzungen zum Initiierungszeitpunkt II (z. B. Kostons et al., 2010; van Loon et al., 2013) als auch für jene zum Initiierungszeitpunkt III und IV (z. B. Dunlosky & Rawson, 2012; Mengelkamp & Bannert, 2010) konnte festgestellt werden, dass eine akkurate Selbsteinschätzung in Zusammenhang mit besseren Performanzwerten steht. Allerdings wird die Bedeutung der Akkuratheit der Selbsteinschätzung zu den verschiedenen Initiierungszeitpunkten für die Performanz selten vergleichend betrachtet (Maki et al., 1994). Darüber hinaus berücksichtigen nur wenige Studien Initiierungszeitpunkt I zur Selbsteinschätzung und dessen Beziehung zur Performanz.

Die Studienergebnisse werden auch von Untersuchungen bestätigt, die sich vorrangig mit den Zusammenhängen zwischen der Akkuratheit der Selbsteinschätzung und der Performanz in anschließenden Tests auseinandersetzen, ohne dabei jedoch die Selbstregulation zu berücksichtigen. Nietfeld et al. (2006) untersuchten im Zeitverlauf eines Semesters die Fähigkeit der Lernenden zur Kalibrierung, also inwiefern die von den Lernenden wahrgenommene Güte der Performanz sich der tatsächlichen anpasst, und wiesen nach, dass die verbesserte Fähigkeit zur Kalibrierung einen signifikant positiven Effekt auf die Performanz im finalen Test ausübt. Dem gegenüber stehen jedoch die Ergebnisse von Glenberg & Epstein (1985), die feststellten, dass eine verbesserte Fähigkeit zur Kalibrierung keinen Einfluss auf die Performanz hat bzw. nicht zu besseren Performanzwerten führt. Zu beachten ist jedoch, dass die von Glenberg & Epstein (1985) referierten Werte darauf schließen lassen, dass die Performanz ohnehin schon recht hoch gewesen ist (Aufgaben wurden zu fast 80 % korrekt beantwortet). Dies könnte auf einen möglicherweise bestehenden Deckeneffekt hinweisen. Auch Begg et al. (1992) geben an, dass eine akkurate Selbsteinschätzung über eine adäquate Wahl zu wiederholender Lernelemente keinen Einfluss auf die Performanz der Lernenden im abschließenden Test hat. Diese Differenz zu den anderen Studienergebnissen lässt sich eventuell auf das methodische Setting zurückführen. In der Studie von Begg et al. (1992) wurden die Lernenden angehalten, Wortpaare, die inhaltlich in keinerlei Zusammenhang zu vorher Gelerntem standen, in einem sehr kurzen Zeitraum (Wortpaare wurden vier Sekunden präsentiert) zu lernen und später zu erinnern. Dieses Lernarrangement entspricht jedoch nicht den im schulischen und universitären Kontext tatsächlich anzutreffenden Lernumgebungen. Daher lassen sich nur bedingt Schlüsse in Bezug auf das selbstregulierte Lernen ziehen.

Der Kritikpunkt der Implementierung von inhaltlich wenig relevanten Aufgaben zur Bestimmung der Akkuratheit der Selbsteinschätzung sowie der Untersuchung potentieller Effekte auf die Performanz lässt sich auch für weitere referierte Studien anführen (Glenberg & Epstein, 1985; Maki et al., 1994; Marshall & Brown, 2006; Thiede et al., 2003; van Loon et al., 2013). Diese Studien sind in ihren Ergebnissen ambivalenter als jene Studien, die einen expliziten inhaltlichen Bezug zur Lernsituation der Lernenden aufweisen und stützen die Annahme positiver Zusammenhänge zwischen Selbsteinschätzung und Performanz nur in Teilen. Studien mit inhaltlichem Bezug zur Lernsituation der Lernenden finden sich eher im universitären (z. B. Coutts & Rogers, 1999; Dunlosky & Rawson, 2012; Mengelkamp & Bannert, 2010; Nietfeld et al., 2006) als im schulischen Kontext (z. B. Hewitt, 2002; Kostons et al., 2012; Narciss et al., 2011; Olina & Sullivan, 2004). Überproportional häufig werden dabei Studierende pädagogisch/psychologischer Studiengänge befragt (z. B. Dunlosky & Rawson, 2012; Mengelkamp &

Bannert, 2010; Nietfeld et al., 2006). Die Konzentration auf wenige Fachbereiche lässt sich auch im schulischen Umfeld beobachten. Hier wurde die Mehrzahl der Studien in naturwissenschaftlichen Fächern (v.a. Biologie) durchgeführt (Kostons et al., 2010, 2012; Snyder et al., 2011).

Neben der bedingt gewährten Situationsvalidität der Untersuchungsergebnisse, bestehen weitere Anhaltspunkte, die an der externen Validität zweifeln lassen. Aufgrund der in Teilen geringen Stichprobengrößen (z. B. Dunlosky & Rawson, 2012; Hewitt, 2002; Kostons et al., 2010; Nietfeld et al., 2006) sowie der nicht nachvollziehbaren Stichprobenbildung, ist auch die Populationsvalidität zu hinterfragen. Die Stichproben beschränken sich zudem tlw. auf einzelne Klassen und werden darüber hinaus noch in verschiedene Treatment-Gruppen eingeteilt (z. B. Kostons et al., 2010), sodass nur wenige Lernende einem Treatment ausgesetzt werden. Überdies wird die Zuordnung der Lernenden auf die Treatment- bzw. Kontrollgruppe(n) in manchen Studien unzureichend beschrieben (z. B. Dunlosky & Rawson, 2012). Beispielsweise wird die Homogenität der Treatment- und Kontrollgruppen bezüglich zu prüfender Einflussvariablen auf die Akkuratheit der Selbsteinschätzung der Lernenden nicht untersucht. Somit können die ermittelten Unterschiede zwischen Treatment- und Kontrollgruppe nicht zweifelsfrei auf die Akkuratheit der Selbsteinschätzung zurückgeführt werden. Weiterhin sind die referierten Studienergebnisse aufgrund der Domänenspezifität der Fähigkeit zur akkuraten Selbsteinschätzung auf andere Fachbereiche nur bedingt übertragbar. Hartwig et al. (2012) versuchten den Nachweis zu erbringen, dass es sich beim Monitoring um eine generelle, fächerübergreifende Fähigkeit handelt. Zwar zeigen sich signifikantere Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der Performanz in anderen Fachbereichen als jenen in der die Selbsteinschätzung vorgenommen wurde, jedoch sind diese nur schwach und die Akkuratheit des Monitoring erklärt lediglich 6.8 % der Varianz der Performanz der Lernenden. Somit sollten Untersuchungen in diesem Kontext auch weiterhin die Domänenspezifität berücksichtigen.

Neben den in den Studien hergestellten Bezügen zu bisher Gelerntem lassen sich die methodischen Settings auch dahingehend unterscheiden, welche konkreten Aufgabenstellungen eingesetzt worden sind, anhand derer die Fähigkeit zur Selbsteinschätzung sowie die Performanz gemessen werden sollte. Die Form der Selbsteinschätzung i. S. der konkreten Aufgabenstellung, die der Selbsteinschätzung zugrunde liegt, wurde als ein Einflussfaktor für die Akkuratheit der Selbsteinschätzung identifiziert (s. Kapitel 2.2). Entsprechend ist zu erwarten, dass diese auch einen Einfluss auf die skizzierten Zusammenhänge ausübt. Die betrachteten Studien lassen sich danach differenzieren, inwiefern nach der Selbsteinschätzung neue oder gleiche Lernaufgaben gewählt werden sollten. Lernende mit akkurater Selbsteinschätzung sind eher in

der Lage, adäquate Lerninhalte zur Wiederholung zu identifizieren (Dunlosky & Rawson, 2012; Thiede et al., 2003; van Loon et al., 2013). Wie erwartet wurde, schneiden die Lernenden mit akkuraterer Selbsteinschätzung in abschließenden Performanztests besser ab. Es zeigt sich, dass dieser Zusammenhang auch für die Wahl neuer Lerngelegenheiten zu beobachten ist (Kostons et al., 2010, 2012).

Zusätzlich zum Kontext der Aufgabe, in der die Selbsteinschätzung vorgenommen wird, ist auch das Instrument, das zur Initiierung der Selbsteinschätzung eingesetzt wird, eine wichtige Komponente, die die tlw. bestehende Ambivalenz in den Studienergebnissen bedingen kann. Die Instrumente unterscheiden sich dabei insbesondere in ihrer Komplexität. Zu den weniger komplexen Instrumenten zählen jene, die Selbsteinschätzungen über binäre Checklisten einfordern (z. B. Hartwig et al., 2012) oder prozentuale Angaben bzw. absolute Angaben zur Richtigkeit der gelösten Aufgaben verlangen (z. B. Mengelkamp & Bannert, 2010; Nietfeld et al., 2006; Snyder et al., 2011). Verbreitet werden auch Likert-Skalen mit unterschiedlichen Ausprägungen eingesetzt (z. B. Begg et al., 1992; Coutts & Rogers, 1999; Glenberg & Epstein, 1985; Maki et al., 1994; Narciss et al., 2011). Zu den komplexeren Instrumenten sind jene zu fassen, in denen Rubrics eingesetzt worden sind (z. B. Hewitt, 2002; Olina & Sullivan, 2004; Panadero & Romero, 2014). Fraglich ist, inwieweit die Fähigkeit zur Selbsteinschätzung mit diesen Instrumenten akkurat gemessen werden kann. Es mangelt zum einen oftmals an einem konkreten Kriterienbezug, um die Qualität der Aufgabenbearbeitung einschätzen zu können. Zum anderen sind die verschiedenen Ausprägungen der eingesetzten Likert-Skalen häufig nicht näher definiert. Es ist daher denkbar, dass die Lernenden andere Kriterien zur Einschätzung der Güte ihres Wissens heranziehen als die Dozierenden, deren Fremdeinschätzung in der Regel als Vergleichsmaß zur Bestimmung der Akkuratheit der Selbsteinschätzung dient. Vor diesem Hintergrund sind die Studienergebnisse zur Akkuratheit der Selbsteinschätzung kritisch zu betrachten.

Trotz der referierten Limitationen der dargestellten Studien, zeichnet sich insgesamt ab, dass sich die vermuteten Zusammenhänge zwischen der Fähigkeit zur Selbsteinschätzung und der Performanz in den Studien bestätigen. Dies gilt jedoch nur bedingt für weitere lernbedeutsame Konstrukte. Nietfeld et al. (2006) berichten zwar, dass die Selbsteinschätzung einen positiven Effekt auf die Selbstwirksamkeit der Lernenden hat, allerdings kommen Narciss et al. (2011) und Marshall & Brown (2006) zum Ergebnis, dass die Tendenz zur Überschätzung des eigenen Wissensstandes in Zusammenhang mit einer größeren Zufriedenheit und positiven Lernemotionen steht. Die Bedeutung akkurater Selbsteinschätzungen für weitere lernbedeutsame Konstrukte ist somit noch nicht abschließend geklärt.

3.3.2 Implikationen für die weitere Forschung

Berücksichtigung des Ausmaßes der Fehleinschätzung

Bei näherer Betrachtung der Forschungsarbeiten zur akkuraten Selbsteinschätzung ist bemerkenswert, dass bisherige Studien das Ausmaß der Fehleinschätzungen und die Auswirkungen auf Selbstregulation und Performanz der Lernenden oftmals nicht im Detail berücksichtigen. Es ist zu vermuten, dass die Beziehung zur Performanz vom Grad der Über-/Unterschätzung abhängt. In bisherigen Studien werden oftmals die Akkuratheiten der Selbsteinschätzung zwischen Treatment- und Kontrollgruppen verglichen (Kostons et al., 2012; Thiede et al., 2003), nicht jedoch das konkrete Ausmaß möglicher Fehleinschätzungen und dessen Beziehung zur Performanz. Hinzu kommt, dass der Vergleich von Treatment- und Kontrollgruppe in Teilen wenig zweckmäßig erscheint. So erhält die Treatment-Gruppe bspw. bei Panadero & Romero (2014) ausformulierte Rubrics zur Selbsteinschätzung, während diese der Kontrollgruppe nicht ausgehändigt werden. Die festgestellten Unterschiede in der Akkuratheit der Selbsteinschätzung und der Performanz lassen sich nicht zweifelsfrei auf die Fähigkeit zur Selbsteinschätzung zurückführen. Denkbar ist auch, dass die im Rahmen der Rubrics vermittelten Erwartungshorizonte das Lernergebnis positiv beeinflusst haben.

Mit einer stärkeren Fokussierung auf das Ausmaß eventuell bestehender Fehleinschätzungen ließe sich auch die Aussagekraft der bisherigen Studienergebnisse verbessern. Bislang weisen die bisherigen Studien einen eher deskriptiven Charakter auf. Es konnte in einer Vielzahl an Studien lediglich nachgewiesen werden, dass Treatments, die eine Selbsteinschätzung von den Lernenden verlangen, zu einer besseren Performanz in Posttests führten. Es erscheint an dieser Stelle plausibel, dass die Einforderung einer Selbsteinschätzung die Funktion eines Prompts einnimmt und allein dadurch unabhängig von der Akkuratheit der Selbsteinschätzung bessere Lernergebnisse induziert werden. Um die Wirkungsmechanismen von Selbsteinschätzung, Selbstregulation und Performanz hinreichender erklären zu können, erscheint es notwendig zu sein, weitere lernbedeutsame Konstrukte zu berücksichtigen. In nur wenigen Studien wurden die von den Lernenden aufgebrauchte Lernzeit (z. B. Dunlosky & Rawson, 2012) sowie die Lernmotivation beachtet. Vor dem Hintergrund der in einigen Studien eingesetzten Lernaufgaben, die nicht in Zusammenhang mit zuvor Gelerntem standen, ist zu vermuten, dass die Motivation der Lernenden variiert und somit auf den Zusammenhang von Selbsteinschätzung und Performanz wirken könnte.

Unklar ist demnach nach wie vor, wie genau die Akkuratheit der Selbsteinschätzung auf die Selbstregulation und die Performanz wirkt. Es lassen sich nur wenige stichhaltige Belege finden, die die angenommenen Beziehungen zwischen akkurater Selbsteinschätzung und Performanz stützen und erklären (Dunlosky & Rawson, 2012). In weiteren Studien sollten die Effekte verschiedener Fehleinschätzungsgrade eingehender betrachtet werden. Denkbar ist, dass Lernende, die zu leichter Unterschätzung neigen, durch einen Effekt des „Überlernens“ von der Fehleinschätzung profitieren (Dunlosky & Rawson, 2012). Ein bestimmtes Ausmaß an Überschätzung könnte sich wiederum bspw. über höhere Selbstwirksamkeitserwartungen positiv auf das Lernen auswirken (s. Abbildung 7). Die Auswirkungen gewisser Ausprägungen der Fehleinschätzungen auf das Lernen können Untersuchungsgegenstand weiterer Forschungsarbeiten sein. Eventuell lässt sich ein Bereich an Fehleinschätzung spezifizieren, der nicht nachteilig für die Performanz wirkt (grün schraffierter Bereich in Abbildung 7).

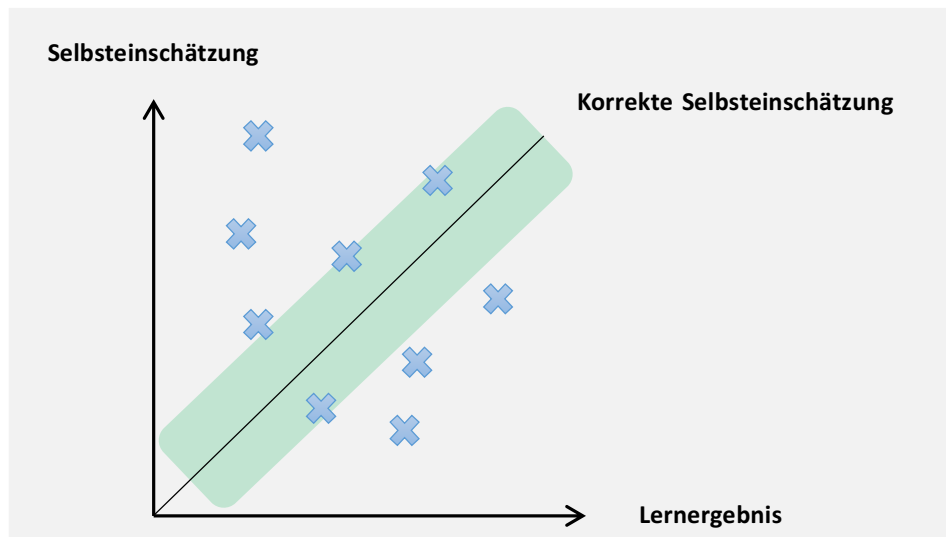


Abbildung 7: Potentielle Zusammenhänge zwischen Fehleinschätzung und Performanz

Inhaltliche Orientierung an tatsächlichen Lerngegebenheiten (Feldstudien)

Da die Fähigkeit zur akkuraten Selbsteinschätzung domänenspezifisch ausgeprägt ist (s. Kapitel 2.2), sollte in zukünftigen Studien eine stärkere inhaltliche Orientierung an den tatsächlichen inhaltlichen Anforderungen der Lernenden berücksichtigt werden. Dies lässt sich durch vermehrten Einsatz von Feldstudien realisieren, womit auch die externe Validität der Studien gesteigert werden könnte. Zudem sollten diese in weiteren Fachbereichen (z. B. im wirtschaftsberuflichen Kontext) durchgeführt werden.

Prüfung der Zusammenhänge unter Berücksichtigung von Aufgaben mit höherem kognitivem Anspruchsniveau

Die in den Studien zur Selbsteinschätzung eingesetzten Lernaufgaben weisen unterschiedliche kognitive Anspruchsniveaus auf, sind jedoch oftmals eher auf niedrigeren Niveaustufen anzusiedeln. In Teilen wurde den Lernenden aufgetragen, einfache Definitionen und Konzepte zu lernen und selbst einzuschätzen (Dunlosky & Rawson, 2012; Thiede et al., 2003; van Loon et al., 2013), weitere Studien fokussierten die Selbsteinschätzung bei der Ausübung von Prozeduren (Coutts & Rogers, 1999; Hewitt, 2002; Olina & Sullivan, 2004; Sitzmann & Johnson, 2012). Ferner werden in manchen Studien Akkuratheit und Performanz hinsichtlich der Selbsteinschätzung mit Einsatz von Aufgaben unterschiedlicher kognitiver Anspruchsniveaus betrachtet (Kostons et al., 2010, 2012; Mengelkamp & Bannert, 2010; Nietfeld et al., 2006; Snyder et al., 2011). Es mangelt jedoch an Studien, die sich mit der Selbsteinschätzung hinsichtlich der Bearbeitung von Aufgaben mit höherem kognitivem Anspruchsniveau auseinandersetzen. Dies inkludiert auch Aufgabenstellungen, die die Selbsteinschätzung der Problemlösekompetenz fordern. Unter Problemlösekompetenz wird die „Fähigkeit, Prozesse kognitiv zu verarbeiten, um Problemsituationen zu verstehen und zu lösen, in denen die Lösungsmethode nicht unmittelbar auf der Hand liegt“ (OECD, 2014) verstanden. Vor dem Hintergrund der gestiegenen Bedeutung der Problemlösekompetenz im Berufsleben (OECD, 2014) sollte sie auch Gegenstand von Untersuchungen im Kontext des selbstgesteuerten Lernens sein.

Entwicklung neuer Instrumente zur Erfassung der Akkuratheit der Selbsteinschätzung

Die Analyse der Studien hat gezeigt, dass die zur Messung der Selbsteinschätzung eingesetzten Instrumente oftmals wenig kriterienbezogen sind und die Niveauabstufungen nicht erläutert werden. Für eine möglichst valide Erfassung der Fähigkeit zur Selbsteinschätzung des Gelerten, erscheint es zweckmäßig, relevante Kriterien zu definieren und den Lernenden zu kommunizieren. Dies schließt ebenso eine deutliche Kenntlichmachung der erreichten Leistungsniveaustufen (z. B. Punkteinteilung bei Rubrics) ein. Zu unterscheiden ist dabei, inwiefern das Instrument allgemein oder inhaltspezifisch ausgerichtet ist. Allgemeinere, auf ein Fachgebiet fokussierte Instrumente böten den Vorteil des effizienteren Einsatzes im universitären Lehrbetrieb, da eine inhaltspezifische Anpassung entfällt. Für eine allgemeinere Ausrichtung könnten verschiedene Taxonomie-Ansätze (z. B. Krathwohl, 2002) hilfreich sein. Diese finden sich bislang nur in wenigen Studien. Alaoutinen (2012) prüfte die Fähigkeit zur akkuraten Selbsteinschätzung von Lernenden in einem Programmierkurs mittels einer taxonomiebasierten Skala, die in Anlehnung an Bloom (1984) entwickelt wurde. Die Lernenden wurden angehalten, für

jede Lerneinheit anzugeben, auf welchem kognitiven Anspruchsniveau ihr erworbenes Wissen einzustufen ist. Die von Alaoutinen (2012) entwickelte Skala ist spezifisch für das Programmieren gedacht, eine Übertragung auf andere Fachgebiete scheint jedoch möglich zu sein, da die Taxonomieebenen nach Bloom kontextunabhängig sind. Barney et al. (2012) entwickelten Rubrics in Anlehnung an die SOLO-Taxonomie von Biggs & Collins (1982) zur Selbsteinschätzung. Der Einsatz solcher kriterienbezogener Instrumente zur Selbsteinschätzung würde es ermöglichen, dass Lernende und Lehrende die gleichen Kriterien für die Beurteilung des Lernergebnisses heranziehen. Somit könnten validere Informationen zur Fähigkeit der Lernenden zur Selbsteinschätzung gewonnen werden.

4 Fragestellung und Hypothesen

Vor dem Hintergrund der in Kapitel 3.3 aufgezeigten Limitationen der bisherigen Forschung zur Selbsteinschätzung im Lernprozess und der damit verbundenen Implikationen wird beabsichtigt, den Zusammenhang zwischen (in-)akkuratere Selbsteinschätzung und Performanz im wirtschaftswissenschaftlichen Kontext zu eruieren (Fragestellung 1). Dieser wird im Rahmen einer regulären universitären Pflichtveranstaltung („Grundlagen des externen Rechnungswesens“) untersucht, indem verschiedene, in Teilen kriterienorientierte Instrumente zur Selbsteinschätzung zu verschiedenen Initiierungszeitpunkten eingesetzt werden. Damit kann auch der Frage nachgegangen werden, mit welcher Form der Selbsteinschätzung bzw. zu welchem Initiierungszeitpunkt eine akkuratere Selbsteinschätzung gefördert werden kann (Fragestellung 2). Weiterhin werden Aufgaben mit unterschiedlichen kognitiven Anspruchsniveaus eingesetzt und entsprechend in den Items zur Selbsteinschätzung berücksichtigt. Das Ausmaß möglicher Fehleinschätzungen könnte dadurch eventuell reduziert werden.

Zur Beantwortung der ersten Fragestellung wird in einem ersten Schritt die Akkuratheit der Selbsteinschätzung von Studierenden im wirtschaftswissenschaftlichen Kontext betrachtet. Bisherige Studien zeigen, dass auch Studierende in Teilen zu Fehleinschätzungen neigen. Cassidy (2007) kommt zum Ergebnis, dass jeder vierte Studierende in der Selbsteinschätzung von der Fremdeinschätzung abweicht. Es wird davon ausgegangen, dass Studierende im wirtschaftswissenschaftlichen Kontext mehrheitlich zu akkuraten Selbsteinschätzungen neigen, jedoch in Teilen auch inakurate Selbsteinschätzungen auftreten (Hypothese 1_a). Das Ausmaß der Fehleinschätzungen sollte sich jedoch zu späteren Initiierungszeitpunkten der Selbsteinschätzung (nach der Aufgabenbearbeitung) und zum späteren Messzeitpunkt (am Ende des Semesters) reduzieren (Hypothese 1_b).

Bisherige Studien kamen überwiegend zu dem Ergebnis, dass eine akkuratere Selbsteinschätzung des eigenen Wissensstandes auch zu besseren Lernergebnissen führt (s. Kapitel 3). Es wird erwartet, dass sich dieser Zusammenhang auch bei Studierenden der Wirtschaftswissenschaften zeigt (Hypothese 2). Darüber hinaus wird jedoch auf Basis der bisher uneinheitlichen Forschung – Studienergebnisse liefern zum einen Hinweise, dass Lernende in Teilen über wenig akkurate Selbsteinschätzungen verfügen, zum anderen wird jedoch von positiven Effekten der Initiierung der Selbsteinschätzung auf das Lernen ausgegangen – vermutet, dass sich auch ein bestimmendes Ausmaß an Fehleinschätzung nicht negativ auf die Performanz zum späteren Messzeitpunkt auswirkt (Hypothese 5).

Eine Erklärung hierfür könnte im unterschiedlichen Einsatz metakognitiver Strategien liegen. Studierende, die zur Unterschätzung neigen, setzen vermutlich mehr Strategien zur Steuerung des weiteren Lernprozesses ein als Studierende mit akkurater Selbsteinschätzung (Hypothese 3). Außerdem dürften sie mehr Zeit mit dem Lernen bereits bekannter Inhalte verbringen als Studierende mit akkurater Selbsteinschätzung (Hypothese 4).

Hinsichtlich der Ausgestaltung des Instrumentes ist zu erwarten, dass die Studierenden sich mit jenen Abschnitten des Instrumentes, die einen stärkeren Kriterienbezug aufweisen, akkurater einschätzen können, als mit den Passagen, die weniger konkrete Einschätzungskriterien beinhalten (Hypothese 6). Nachfolgend findet sich eine tabellarische Übersicht der betrachteten Fragestellungen und Hypothesen.

F₁	Welcher Zusammenhang besteht zwischen der Akkuratheit der Selbsteinschätzung und dem Lernergebnis im wirtschaftswissenschaftlichen Kontext?
H _{1a}	Studierende im wirtschaftswissenschaftlichen Kontext weisen überwiegend akkurate Selbsteinschätzungen auf, d.h. die Studierenden weichen in ihren Selbsteinschätzungen nicht mehr als einen Skalenpunkt von der tatsächlichen Performanz ab.
H _{1b}	Es wird erwartet, dass die Akkuratheit der Selbsteinschätzung zum zweiten Initiierungs- und Messzeitpunkt zunimmt (Maki et al., 1994, Kostons et al., 2012).
H ₂	Eine akkuratere Selbsteinschätzung steht im wirtschaftswissenschaftlichen Kontext in Zusammenhang mit einem besseren Lernergebnis (Dunlosky & Rawson, 2012; Nietfeld et al., 2006; Thiede et al., 2003).
H ₃	Studierende, die zur Unterschätzung neigen, setzen mehr Strategien zur Steuerung des weiteren Lernprozesses (Lernstrategien) ein als Studierende mit akkurater Selbsteinschätzung (Thiede et al., 2003).
H ₄	Zur Unterschätzung tendierende Studierende verbringen mehr Zeit mit dem Lernen der Inhalte als Studierende mit akkurater Selbsteinschätzung (Dunlosky & Rawson, 2012).
H ₅	Ein gewisses Ausmaß an Fehleinschätzungen ist für die Performanz der Studierenden zum späteren Messzeitpunkt effektneutral (Coutts & Rogers, 1999; Sitzmann & Johnson, 2012).
F₂	Welche Zusammenhänge bestehen zwischen dem Einsatz eines stärker kriterienbezogenen Instruments zur Selbsteinschätzung und der Akkuratheit der Selbsteinschätzung?
H ₆	Studierende können sich mit jenen Abschnitten des Instrumentes, die einen stärkeren Kriterienbezug aufweisen, akkurater einschätzen, als mit jenen Abschnitten, die weniger konkrete Einschätzungskriterien beinhalten.

Tabelle 4: Übersicht der Fragestellungen und Hypothesen

5 Untersuchungsdesign

5.1 Anlage der Untersuchung

Um die Fähigkeit der Studierenden zur akkuraten Selbsteinschätzung und die Zusammenhänge dieser mit der Performanz über die aufgezeigten Initiierungszeitpunkte (s. Kapitel 2.1) hinweg zu bestimmen und zu vergleichen, werden Studierende verschiedener Fachrichtungen im Semesterverlauf befragt und um eine Selbsteinschätzung hinsichtlich der Bearbeitung lernzielbezogener Aufgaben in der Veranstaltung „Grundlagen des externen Rechnungswesens“ gebeten. Dieses Fach ist auch in Studiengängen, die Wirtschaftswissenschaften nicht als Hauptfach umfassen, von Bedeutung, da es in den Prüfungsordnungen verankert ist. Eine Analyse der Modulkataloge in diesen Studienfächern zeigt, dass bspw. Studierende der Wirtschaftspädagogik etwa ein Viertel ihrer gesamten ECTS-Punkte in der Betriebswirtschaftslehre im Rechnungswesen erwerben. Ferner erlaubt die im Rahmen des Faches vorgenommene Darstellung grundlegender Geld- und Leistungsströme einer Unternehmung einen umfassenden Einblick in betriebswirtschaftliche Leistungsprozesse und ist somit von besonderer Bedeutung für das Verständnis betriebswirtschaftlicher Prozesse. Aus diesen Gründen wird die Fähigkeit zur Selbsteinschätzung der Studierenden im Rahmen der Veranstaltung „Grundlagen des externen Rechnungswesens“ betrachtet.

Die erste Datenerhebung (Messzeitpunkt I) erfolgt in der sechsten Vorlesungswoche im Rahmen von zwei Übungsterminen zur Veranstaltung „Grundlagen des externen Rechnungswesens“, die den gleichen Inhalt zum Gegenstand haben. Die Studierenden müssen also nur einen der beiden Termine wahrnehmen. Zu diesem Zeitpunkt haben die Studierenden einige Vorlesungen und Tutorien sowie eine Übung besucht. Damit kann eine erste Testung der bisher in der Veranstaltung angesprochenen Inhalte erfolgen. Eine solche ist Voraussetzung dafür, dass die Akkuratheit einer Selbsteinschätzung geprüft werden kann. Selbsteinschätzungen von den Studierenden werden mit je gleichen Items vor (Initiierungszeitpunkt I²⁴) und nach Bearbeitung der Testaufgaben (Initiierungszeitpunkt II) erfragt. Zusätzlich wird direkt im Anschluss einer jeden Aufgabenstellung von den Studierenden eine Einschätzung vorgenommen, inwieweit sie überzeugt sind, die Aufgaben korrekt bearbeitet zu haben (Initiierungszeitpunkt III²⁵). Dabei werden die eingesetzten Items zur Selbsteinschätzung und die implementierten Testaufgaben

²⁴ In der bisherigen Darstellung entsprach der Initiierungszeitpunkt I Selbsteinschätzungen i. S. von EoL-Urteilen (s. Kapitel 3). Da im Rahmen der Untersuchung jedoch auf dieses Urteil verzichtet wurde, wird die Notation angepasst: Der Initiierungszeitpunkt I stellt hier eine Selbsteinschätzung i. S. des JoL dar.

²⁵ Da die Notation der Initiierungszeitpunkte nicht der chronologischen Abfolge in den beiden Befragungen entspricht, wird in der Folge der dritte Initiierungszeitpunkt jeweils mit *Überzeugungen* umschrieben.

auf verschiedene kognitive Anspruchsniveaus ausgerichtet sein. Das wesentliche Unterscheidungskriterium bei der Erstellung der Items zur Selbsteinschätzung und der Auswahl der Testaufgaben ist dabei die jeweils angesprochene Wissensart. Ferner werden bei der Konstruktion der Befragung unterschiedliche Aufgabenformate berücksichtigt (z. B. offene Fragestellungen, MC-Aufgaben und die Bildung von Buchungssätzen). Zusätzlich sollten die Studierenden die bislang eingesetzten Lernstrategien und den Zeitaufwand für das Lernen der Veranstaltungsinhalte angeben.

Die Befragung wird in der zwölften Vorlesungswoche wiederholt (Messzeitpunkt II), um dem gegen Semesterende vermutlich fortgeschrittenen Lernprozess seitens der Studierenden Rechnung zu tragen. Damit die Vergleichbarkeit der Ergebnisse gewährleistet ist, werden die eingesetzten Aufgaben und Items zwischen den Messzeitpunkten nicht verändert.

Auf eine Selbsteinschätzung zu Semesterbeginn im Rahmen eines EoL-Urteils wird in dieser Untersuchung verzichtet. Eine Bestimmung der Akkuratheit solcher Urteile ist nur schwer möglich, da ohne jegliches Wissen der Lernenden über Inhalte des externen Rechnungswesens eine Performanztestung, die zur Bestimmung der Akkuratheit der Selbsteinschätzung benötigt wird, von geringer Aussagekraft sein dürfte. Darüber hinaus ist mit dem Verzicht auf die Erfassung eines EoL-Urteils zu Beginn des Semesters die Intention verbunden, die Teilnahmebereitschaft der Studierenden durch eine weitere Erhöhung der Itemzahl nicht zu beeinträchtigen. In Abbildung 8 werden die Inhalte der Befragung für beide Messzeitpunkte zusammenfassend dargestellt.

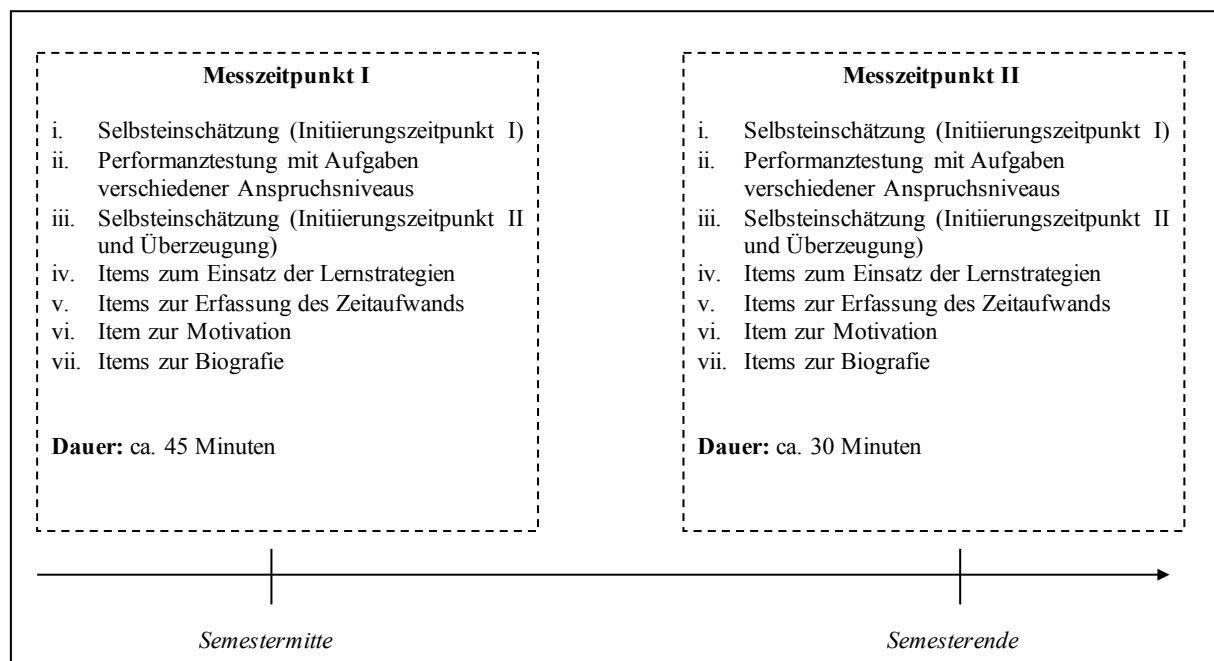


Abbildung 8: Anlage der Untersuchung

5.2 Designkomponenten des ersten Messzeitpunktes

5.2.1 Stichprobe

Im Rahmen der ersten Erhebung in der sechsten Vorlesungswoche konnte ein Rücklauf von 392 Fragebögen registriert werden. Zum ersten Übungstermin konnten 292 bearbeitete Fragebögen für die Auswertung herangezogen werden, zum zweiten Übungstermin weitere 97 Fragebögen. Drei Fragebögen waren für eine Auswertung nicht hinreichend bearbeitet²⁶ und wurden somit ausgeschlossen. Von den insgesamt 389 Studierenden sind 67.9 % weiblich und 26.0 % männlich. Bei weiteren 6.1 % der Studierenden fehlt die Angabe zum Geschlecht. Die Studierenden sind im Durchschnitt 21.01 Jahre alt ($SD = 2.40$ Jahre) und mehrheitlich in den Bachelorstudiengängen *Kultur und Wirtschaft* (32.4 %) sowie *Wirtschaftspädagogik* (28.8 %) eingeschrieben. Ihre Hochschulzugangsberechtigung (HZB) erwarben die Studierenden überwiegend am *allgemeinbildenden Gymnasium* (72 %). Lediglich 14.1 % der Probanden besuchte zuvor ein *Wirtschaftsgymnasium*. Eine *kaufmännische Berufsausbildung* vor dem Studium haben 8.7 % der befragten Studierenden abgeschlossen. Somit ist davon auszugehen, dass die teilnehmenden Studierenden zum ersten Messzeitpunkt über ein geringes Vorwissen im Bereich des externen Rechnungswesens verfügen. Diese Annahme wird gestützt von der Tatsache, dass lediglich 18.3 % der Studierenden angaben, sich bereits im schulischen Kontext mit Inhalten des externen Rechnungswesens auseinandergesetzt zu haben. In Tabelle 5 sind die demografischen Angaben der teilnehmenden Studierenden zum ersten Messzeitpunkt aufgeschlüsselt nach Übungstermin zusammengefasst.

²⁶ Die Bearbeitung konzentrierte sich nur auf die Angabe des Codes. Es wurden keine Selbsteinschätzungen vorgenommen bzw. Testaufgaben bearbeitet.

Kriterium		Übungs- termin I	Übungs- termin II	Σ
N		292	97	389
Geschlecht	weiblich	207	57	264
	männlich	67	34	101
	fehlende Angabe	18	6	24
Alter	M	20.95	21.20	21.01
	SD	2.08	3.17	2.40
	fehlende Angabe	28	7	35
Studienfach	Bachelor Kultur und Wirtschaft	76	50	126
	Bachelor Unternehmensjurist	32	7	39
	Bachelor VWL	18	15	33
	Bachelor Wirtschaftsinformatik	3	4	7
	Bachelor Wirtschaftspädagogik	111	1	112
	Lehramt an Gymnasien	4	6	10
	Master Psychologie	1	1	2
	sonstiges	28	6	34
	fehlende Angabe	19	4	23
HZB	allgemeinbildendes Gymnasium	204	76	280
	Wirtschaftsgymnasium	45	10	55
	anderer Schultyp	26	8	34
	fehlende Angabe	17	3	20
abgeschlossene kaufmännische Berufsausbildung	ja	28	6	34
	nein	246	88	334
	fehlende Angabe	18	3	21
schulische Erfahrungen mit Rechnungswesen	ja	56	15	71
	nein	218	80	298
	fehlende Angabe	18	2	20

Tabelle 5: Charakterisierung der Stichprobe zu Messzeitpunkt I

5.2.2 Instrument

5.2.2.1 Darstellung und Klassifizierung der Testaufgaben

Die Akkuratheit der Selbsteinschätzung wird auf Basis eines Abgleichs der Performanz der Studierenden bei der Bearbeitung von Testaufgaben und der Selbsteinschätzung bzgl. des Abschneidens bei ebendiesen bestimmt. Hierzu werden Studierende befragt, die die Veranstaltung „Grundlagen des externen Rechnungswesens“ besuchen. In dieser Veranstaltung haben sie die Möglichkeit, eine Vielzahl an Übungsaufgaben sowohl während der Übungen und Tutorien als auch im Selbststudium zu bearbeiten. Die in der Veranstaltung sowie der veranstaltungsbegleitenden Literatur (Wüstemann, 2015) eingesetzten Aufgaben sind nach drei Aufgabenformaten geordnet. Es werden einerseits Textaufgaben eingesetzt, mit denen zumeist das basale Verständnis über bestimmte Inhalte des externen Rechnungswesens erfragt werden soll (z. B. *Nennen und erläutern Sie die in § 1 des Umsatzsteuergesetzes genannten Grundtatbestände steuerbarer Umsätze*). Andererseits bilden Fallstudien, die den Charakter von Transferaufgaben aufweisen, einen Schwerpunkt. Durch diese werden konkrete Geschäftssituationen bearbeitet (z. B. *Kontenabschluss und Ermittlung des Geschäftsjahresergebnisses*). Zuletzt wird in einer Vielzahl von Aufgaben die Bildung von Buchungssätzen gefordert (z. B. *Am Jahresende gewährt die M-OHG einem Kunden einen Bonus i. H. v. 2 % auf dessen Jahresumsatz aus Wareneinkäufen i. H. v. 390.000 Euro (brutto). Der Bonus wird mit noch bestehenden Forderungen gegenüber dem Kunden verrechnet*).

Diese drei Aufgabenformate finden sich für fast alle 14 in der Veranstaltung behandelten Gliederungspunkte²⁷ wieder. Für die Konzipierung der Befragung zum ersten Messzeitpunkt ist vorgesehen, dass Inhalte aus jenen Gliederungspunkten ausgewählt werden, die zu diesem Zeitpunkt bereits in Vorlesung, Übung oder Tutorium thematisiert worden sind. Darüber hinaus soll bei der Auswahl der Testaufgaben sichergestellt werden, dass die darin abgedeckten Inhalte auch für die Abschlussprüfung von Relevanz sind. Laut Angaben aus dem Modulkatalog (Universität Mannheim, 2017) sollen die Studierenden „fundierte Kenntnisse der Buchführung und Buchführungstechnik sowie ein grundlegendes Verständnis der Grundsätze ordnungsmäßiger Buchführung (GoB)“ erwerben. Dies beinhaltet insbesondere die Verbuchung von Geschäftsvorfällen. Aus diesem Grund sollen Textaufgaben, MC-Aufgaben und Fallstudien aus

²⁷ Die Kursinhalte erstrecken sich auf die Themen Inventar und Bilanz, Erfolgsneutrale Geschäftsvorfälle, Erfolgswirksame Geschäftsvorfälle, Warenverkehr, Umsatzsteuer, Anschaffungskosten, Umsatzerlöse und Zahlungsverkehr, Personalaufwand, Abschreibungen auf das Anlagevermögen, Abschreibungen auf das Umlaufvermögen, Rechnungsabgrenzungsposten, Rückstellungen, Herstellungskosten sowie Ergebnisrechnung und Erfolgsverbuchung bei ausgewählten Rechtsformen.

den Bereichen *erfolgsneutrale und erfolgswirksame Geschäftsvorfälle* sowie der *Umsatzsteuer* näher betrachtet werden.

	Item	Aufgabenstellung
Textaufgaben	TA131	Was versteht man unter dem <i>Prinzip der getrennten Kontenführung</i> ?
	TA132	Nennen Sie die in § 1 UStG genannten Tatbestandselemente steuerbarer Umsätze.
	TA133	Was versteht man unter dem <i>Belegprinzip</i> ? Gehen Sie auf die Anforderungen einer ordnungsmäßigen Belegorganisation ein.
MC-Aufgaben		<p>Im Folgenden werden Ihnen einige Buchungssätze geschildert und mögliche zugrundeliegende Sachverhalte aufgeführt. Markieren Sie jeweils die möglicherweise zugrundeliegenden Sachverhalte. Trifft kein Sachverhalt zu, markieren Sie dies entsprechend. Beachten Sie, dass Mehrfachantworten möglich sind.</p> <p>Buchungssätze:</p> <p>MC141 Andere Forderung an Zinsertrag MC142 Privatentnahme an Bank MC143 Rohstoffaufwand an Rohstoffe MC144 GuV-Konto an Eigenkapital MC145 Kurzfristige Verbindlichkeiten an langfristige Verbindlichkeiten MC146 Fuhrpark an Privateinlage MC147 Mietaufwand an Bank</p> <p>mögliche Sachverhalte:</p> <ul style="list-style-type: none"> ○ Erfolgsneutrale Buchung ○ Aufwand/Ertrag ○ Kontenabschluss ○ Aktivtausch ○ Passivtausch ○ Kein Sachverhalt trifft zu
Buchungsaufgaben		Bilden Sie für die nachfolgenden Sachverhalte den entsprechenden Buchungssatz! Die Umsatzsteuer ist dabei zunächst zu vernachlässigen.
	BS161	Auf dem Bankkonto der M-OHG gehen 450 € für eine an den Unternehmer Z vermietete Lagerhalle ein.
	BS162	Tilgung eines Bankkredits: Neben der Tilgungsrate i. H. v. 3.000 € sind die fälligen Zinsen i. H. v. 120 € von der M-OHG zu überweisen.
	BS163	Am Geschäftsjahresende ist bei der M-OHG das Konto „Privatentnahmekonto“ abzuschließen. Die Summe der Privatentnahmen der laufenden Geschäftsperiode beträgt insgesamt 25.900 €.
		Bitte stellen Sie für die folgenden Sachverhalte die Buchungssätze auf und entscheiden Sie selbst, ob jeweils die <i>Umsatzsteuer</i> zu verrechnen ist. Rechnen Sie dabei mit einem <i>Umsatzsteuersatz</i> von 10 %.
	BS164	Um den Ablauf beim Postversand zu beschleunigen, kauft die M-OHG eine Frankiermaschine für 260 € (netto) auf Ziel.
	BS165	Unternehmer M legt 4.500 € in die betriebliche Kasse der M-OHG ein.
	BS166	Verkauf von Waren im Gesamtwert von 55.000 € (brutto) auf Ziel.

	Item	Aufgabenstellung
Fallstudie (bzw. Transferaufgabe)	FS181	Ermitteln Sie den Gewinn und das Eigenkapital der M-OHG für das Geschäftsjahr t_1 anhand der folgenden Angaben: Anlagevermögen am 31.12. t_1 : 350.000 € Umlaufvermögen am 31.12. t_1 : 250.000 € Verschuldungsgrad am 31.12. t_1 : 4 Privatentnahmen während des Geschäftsjahres t_1 : 2.500 € monatlich Privateinlage am 30.09. t_1 : 20.000 € Privateinlage am 15.03. t_1 : 5.000 € Eigenkapital am 31.12. t_0 : 100.000 € <i>Hinweis: Der Verschuldungsgrad ist definiert als Fremdkapital/Eigenkapital</i>

Tabelle 6: Übersicht der für die Befragung ausgewählten Aufgaben

In Tabelle 6 sind die ausgewählten Aufgaben dargestellt, die der veranstaltungsbegleitenden Literatur unverändert entnommen wurden. Insgesamt müssen die Studierenden drei Textaufgaben, sieben MC-Aufgaben, sechs Buchungsaufgaben und eine Fallstudie²⁸ bearbeiten.

Das Ziel der Arbeit besteht u. a. darin, die Akkuratheit der Selbsteinschätzung für die dargestellten Aufgaben mit unterschiedlichen kognitiven Anspruchsniveaus zu bestimmen. Um dem gerecht werden zu können, müssen die im Rahmen der Befragung eingesetzten Aufgaben zunächst dahingehend geprüft werden, ob mit ihnen auch verschiedene kognitive Anspruchsniveaus angesprochen werden. Hierzu werden im Folgenden in einem Exkurs zunächst einige in der bisherigen Forschung genutzten Ansätze zur Klassifizierung von Aufgaben dargestellt und hinsichtlich der Relevanz und Anwendbarkeit für dieses Forschungsvorhaben gewürdigt.

Exkurs: Klassifizierung von Aufgaben

Eine verbreitete und in der Forschung häufig eingesetzte Möglichkeit, die Aufgaben hinsichtlich ihres kognitiven Anspruchsniveaus zu klassifizieren, besteht in der Anwendung der revidierten Form der Taxonomie nach Bloom (Krathwohl, 2002). Die ursprüngliche Taxonomie nach Bloom, die diesem Ansatz zugrunde liegt, unterscheidet sechs Kategorien, in die Lernziele und Lernaufgaben zugeordnet werden können: *Wissen*, *Verständnis*, *Anwendung*, *Analyse*, *Synthese* und *Bewertung*. Mit Ausnahme der Kategorie *Anwendung* sind die Taxonomien in weitere Subkategorien zergliedert, wobei die Lernziele vom Konkreten zum Abstrakten angeordnet und

²⁸ Die Bezeichnung der Aufgaben wurden der vorlesungsbegleitenden Literatur (Wüstemann, 2015) unverändert übernommen, obwohl die skizzierte Aufgabe die Kriterien einer Fallstudie wie bspw. Offenheit und Vielschichtigkeit der Fragestellung (Preckel, 2006) nicht erfüllt. In der Folge wird die Aufgabe FS181 daher als Transferaufgabe bezeichnet.

hierarchisch aufgebaut sind. Diese Kategorien wurden in der revidierten Fassung um den Aspekt der Wissensdimension erweitert. Unterschieden werden dabei nach Krathwohl (2002) *Faktenwissen*, *konzeptuelles Wissen*, *prozedurales Wissen* und *metakognitives Wissen*. Das Faktenwissen beinhaltet nach Anderson & Krathwohl (2001, S. 46) die basalen Elemente (z. B. Terminologien und Details) die ein Lerner kennen muss, um ein Problem lösen zu können. Das konzeptuelle Wissen unterscheidet sich vom Faktenwissen dahingehend, dass Zusammenhänge (z. B. Prinzipien, Theorien und Modelle) zwischen den basalen Elementen verstanden werden müssen. Die basalen Elemente sind im Gegensatz zum Faktenwissen nicht mehr isoliert zu betrachten. Das prozedurale Wissen umfasst das Wissen über die Anwendung bestimmter Fähigkeiten, Techniken und Prozeduren, um eine Aufgabe bzw. ein Problem zu lösen. Mit metakognitivem Wissen wird das Wissen über die eigenen Kognitionen (z. B. Strategiewissen, Wissen über kognitive Anforderungen einer Aufgabe) beschrieben.

Die Wissensdimensionen werden den kognitiven Prozessdimensionen, die bei der Bearbeitung einer Aufgabe abgerufen werden, gegenübergestellt. Mit dem daraus resultierenden Raster (s. Abbildung 9) können Aufgaben hinsichtlich ihres kognitiven Anspruchsniveaus klassifiziert werden.

Cognitive Process Dimension Knowledge Dimension	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Metacognitive Knowledge						

Abbildung 9: Revidierte Taxonomien kognitiver Prozesse nach Anderson & Krathwohl (2001)

Die Bloom'sche Taxonomie, die seit den 1970er-Jahren Verbreitung gefunden hat, ist in Teilen jedoch auch kritisch zu sehen (Maier, Kleinknecht, & Metz, 2010). Insbesondere wird die fehlende Trennschärfe der Taxonomiestufen bemängelt (Kleinknecht, Maier, Metz, & Bohl, 2011), was dazu führt, dass in vielen Studien, in denen auf die Taxonomie von Bloom und Anderson zurückgegriffen wird, zu einer zweifaktoriellen Struktur übergegangen und lediglich zwischen Low-Level- und High-Level-Aufgaben unterschieden wird (Maier et al., 2010, S. 29).

Daran anknüpfend verfolgten Maier et al. (2010) das Ziel, ein allgemeindidaktisches Kategoriensystem zu entwickeln, welches eine Analyse des kognitiven Aktivierungspotentials einer Aufgabe ermöglicht. Unterschieden werden dabei sieben Kategorien:

(i) *Art des Wissens*

Maier et al. (2010) greifen bei der Beschreibung der Wissensarten auf die Unterscheidung von Anderson & Krathwohl (2001) zurück. Demnach wird zwischen Faktenwissen, konzeptuellem Wissen, prozeduralem Wissen und metakognitivem Wissen differenziert.

(ii) *Kognitive Prozesse*

Bei der Abstufung der kognitiven Prozesse wird versucht, der fehlenden Trennschärfe der Bloom'schen Taxonomie Rechnung zu tragen. Maier et al. (2010) unterscheiden zunächst Reproduktions- und Transferaufgaben. *Reproduktionsaufgaben* erfordern eine Erinnerungsleistung, d. h. Wissen muss in der Form abgerufen werden, wie es auch im Langzeitgedächtnis gespeichert ist. Ein *naher Transfer* liegt nach Maier et al. (2010, S. 33) dann vor, wenn „sich die Aufgabensituation geringfügig von bereits bekannten oder geübten Aufgaben bzw. von der Lernsituation unterscheidet“. Dies könnte bspw. vorliegen, wenn Klarheit über das in der Aufgabensituation anzuwendende Wissen besteht. Bei einem *weiten Transfer* hingegen muss das Wissen in einer neuen und unbekannten Situation angewandt werden. Darüber hinaus besteht keine Klarheit darüber, welches Wissen in dieser Situation gebraucht wird. Grundsätzlich ist dieses Wissen jedoch vorhanden. Hierin unterscheidet sich der *weite Transfer* vom *kreativen Problemlösen*, bei welchem vom Lernenden eine kreative Leistung erfordert wird, sodass neues Wissen entsteht.

(iii) *Anzahl der Wissenseinheiten*

Die Zahl der zur Aufgabenbewältigung zu aktivierenden Wissenseinheiten determiniert nach Maier et al. (2010) auch die Komplexität einer Aufgabe. Eine Wissenseinheit ist zu aktivieren, wenn bspw. nach einer Bezeichnung gefragt wird oder ein Fachbegriff erklärt werden soll. Bis zu vier Wissenseinheiten sind bspw. bei Rechenoperationen zu berücksichtigen. Mehr als vier Wissenseinheiten erfordern bspw. die Wiedergabe mehrerer Termini. Aufgaben sind in der Regel dann komplexer, wenn die Wissenseinheiten zu kombinieren sind und nicht unabhängig voneinander bestehen.

(iv) *Offenheit der Aufgabenstellung*

Hierbei wird unterschieden, inwiefern Anfangs- und Zielzustand einer Aufgabensituation sowie der sich dazwischen befindliche Transformationsprozess eindeutig oder offen sind. Ist der Arbeitsauftrag eindeutig definiert und genau eine Lösung gesucht, sind die Aufgaben definiert und konvergent. Sind die Aufgaben eindeutig formuliert, jedoch

mehrere Lösungen bzw. Lösungswege möglich, handelt es sich um definierte und divergente Aufgaben. Schlecht definierte und divergente Aufgaben weisen weder eine klare Aufgabenstellung noch einen eindeutigen Lösungsweg auf.

(v) *Lebensweltbezug*

Maier et al. (2010) differenzieren zwischen Aufgaben ohne Lebensweltbezug, Aufgaben mit konstruiertem Lebensweltbezug, Aufgaben mit konstruiertem, aber authentisch wirkendem Lebensweltbezug und Aufgaben mit realem Lebensweltbezug.

(vi) *Sprachlogische Komplexität*

Es werden Aufgaben mit einfacher, mittlerer und hoher sprachlicher Komplexität unterschieden.

(vii) *Repräsentationsformen des Wissens*

Aufgaben können sich zunächst *innerhalb einer Repräsentationsform* bewegen. Daneben kann mit Aufgaben gefordert werden, dass zur Lösung *verschiedene Repräsentationsformen integriert* werden müssen. Eine solche Aufgabe liegt vor, wenn diese Wissen in verschiedenen Repräsentationsformen verlangt, die zur Lösung integriert werden müssen. Bei *Integration und Transformation des Wissens* muss der Lernende „für die Aufgabenlösung das vorliegende Wissen in eine Repräsentationsform transformieren, die nicht durch die Aufgabe vorgegeben wird“ (Maier et al., 2010, S. 37).

Die nachfolgende Abbildung 10 fasst die Kategorien nach Maier et al. (2010) zusammen.

Kategorien	Ausprägungen			
Art des Wissens	Faktenwissen	Konzeptuelles Wissen	Prozedurales Wissen	Metakognitives Wissen
Kognitive Prozesse	Reproduktionsaufgaben	Naher Transfer	Weiter Transfer	Kreatives Problemlösen
Anzahl der Wissenseinheiten	1 Wissenseinheit	1-4 Wissenseinheiten	Mehr als 4 Wissenseinheiten	
Offenheit der Aufgabenstellung	Definiert und konvergent	Definiert und divergent	Schlecht definiert und divergent	
Lebensweltbezug	Ohne Lebensweltbezug	Mit konstruiertem Lebensweltbezug	Mit konstruiertem, aber authentisch wirkenden Lebensweltbezug	Mit realem Lebensweltbezug
Sprachlogische Komplexität	Niedrig	Mittel	Hoch	
Repräsentationsformen des Wissens	1 Repräsentationsform	Integration verschiedener Repräsentationsformen	Integration und Transformation des Wissens	

Abbildung 10: Kategoriensystem zur Analyse von Aufgaben (Maier et al., 2010)

Die Anwendbarkeit der Kategorien von Maier et al. (2010) auf die Aufgaben, die im Rahmen der Veranstaltung „Grundlagen des externen Rechnungswesens“ bearbeitet werden, erscheint in vielen Fällen fraglich und eine adäquate Stufung der Aufgaben nach ihrem kognitiven Anspruchsniveau nur schwer möglich. Die *sprachlogische Komplexität* unterscheidet sich in Aufgabenstellungen des externen Rechnungswesens unwesentlich und ist daher nicht geeignet, das Niveau einer Aufgabe zu determinieren. Auch der *Lebensweltbezug* ist bedingt durch die Relevanz der Inhalte für spätere Tätigkeiten im Rechnungswesen durchgängig gegeben. Ferner erfordern fast alle Aufgaben eine höhere *Repräsentationsform* des Wissens (z. B. Bildung eines Buchungssatzes auf Basis eines gegebenen Geschäftsvorfalles). Die *Anzahl der Wissenseinheiten* ist auf der einen Seite schwierig zu quantifizieren. Auf der anderen Seite beschreiben Maier et al. (2010) zwar, dass die Komplexität einer Aufgabe dann ansteigt, wenn verschiedene Wissenseinheiten zur Lösung einer Aufgabe kombiniert werden müssen, jedoch ist unklar, welche Kombinationen von Wissenseinheiten welche Konsequenzen für das kognitive Anspruchsniveau einer Aufgabe mit sich bringen. Nicht zuletzt erscheint eine Unterscheidung zwischen *nahe* und *weitem Transfer* wenig trennscharf zu sein. So ist es schwierig zu beurteilen, wann eine Aufgabe nur „geringfügig“ von bereits bekannten Aufgaben abweicht. Vor dem Hinter-

grund der in Tabelle 6 skizzierten Aufgaben erscheint es sinnvoller, eine allgemeinere Unterscheidung von Reproduktions- und Transferaufgaben vorzunehmen. Die Fallstudie zur Ermittlung des Gewinns und Eigenkapitals ist dabei als Transferaufgabe zu klassifizieren.

Einige der bereits genannten Elemente allgemeindidaktischer Formen zur Bestimmung von Aufgabenkomplexitäten finden sich auch in den wirtschaftsberuflichen Ansätzen zur Klassifikation von Aufgaben wieder. 2003 wurde auf Basis einer Klassifikationsmatrix ein Fachtest für Wirtschaft und Verwaltung für das Projekt „Untersuchungen der Leistungen, Motivation und Einstellungen II“ (ULME II) erstellt. Hofmeister (2005) erläutert, wie die Klassifikationsmatrix entwickelt worden ist. Den Ausgangspunkt bildete die revidierte Taxonomie nach Bloom von Anderson & Krathwohl (2001). Mit dem Test wurde das Ziel verfolgt, das Leistungsniveau der Lernenden zu einem Zeitpunkt zu eruieren (Hofmeister, 2005). Aus diesem Grund wurde darauf verzichtet, das metakognitive Wissen in der Klassifikationsmatrix zu berücksichtigen. Die Ausdifferenzierung der drei verbliebenen Wissensarten wurde leicht verändert. Das Faktenwissen wird nach *terminologischem* und *empirischem Faktenwissen* getrennt. Unter *terminologischem Faktenwissen* werden Aussagen über „spezifische verbale oder nonverbale Bezeichnungen und Symbole“ gefasst (Hofmeister, 2005, S. 6). Solche Aussagen können Definitionen sein, die ohne Bezüge zu anderen Phänomen auskommen (z. B. Bedeutung von Namen & Symbolen). *Empirisches Faktenwissen* beschreibt, „welche Phänomene, Sachverhalte (etwas gibt es) oder Normen (etwas soll sein) es in der Realität gibt“ (Hofmeister, 2005, S. 8), jedoch werden noch keine Aussagen über Zusammenhänge getroffen. Das *Konzeptwissen* wird nach *Kategorien und Klassifikationen* sowie *Prinzipien und Verallgemeinerungen* unterteilt. Dabei sind *Kategorien und Klassifikationen* „Hilfsmittel, um Terminologien eines Sachgebietes zu strukturieren und übersichtlich zu halten“ (Hofmeister, 2005, S. 9). Demgegenüber beschreiben *Prinzipien und Verallgemeinerungen* „Prozesse und Zusammenhänge zwischen empirischen Phänomenen“ (Hofmeister, 2005, S. 10). Hierunter fällt auch das Wissen über Theorien, Modelle und Strukturen, also das Wissen über Ursache- und Wirkungsbeziehungen. Das prozedurale Wissen wird gegliedert nach *Fertigkeiten und Verfahren* sowie *Techniken und Methoden*. Bei ersterem sind der angestrebte Soll-Zustand und der Bearbeitungsweg zur Erreichung dessen bekannt. Erfordert eine Aufgabe prozedurales Wissen über *Techniken und Methoden*, sind der Zielzustand und die Handlungsschritte dorthin jedoch unsicher und müssen ausgewählt werden.

Die vorgenommenen Unterscheidungen innerhalb der verschiedenen Wissensarten erscheinen für den Kontext des externen Rechnungswesens im Hinblick auf das kognitive Anspruchsniveau von geringer Bedeutung zu sein. So ist bspw. nicht ersichtlich, inwiefern sich *terminolo-*

gisches Faktenwissen von *empirischem Faktenwissen* im kognitiven Anspruchsniveau unterscheiden. Aus diesem Grund werden die Ausdifferenzierungen der Wissensarten nach Hofmeister (2005) nicht für die Klassifizierung der in der Befragung eingesetzten Aufgaben herangezogen.

Neben den Wissensdimensionen wurden auch die kognitiven Leistungsdimensionen nach Bloom in Anlehnung an Metzger, Waibel, Henning, Hodel, & Luzi (1993) verändert. Metzger et al. (1993) unterscheiden drei Kategorien der kognitiven Leistungsdimension: *Informationserinnerung*, *Informationsverarbeitung* und *Informationserzeugung*. Wichtigstes Unterscheidungskriterium hierbei ist das Ausmaß des für die Aufgabenbearbeitung geforderten eigenständigen kognitiven Beitrags. Bei Aufgabenstellungen, die auf *Informationserinnerung* abzielen, ist dieser noch gering, da gelernte Informationen in bekannten Aufgabensituationen unverändert angewandt werden müssen. Aufgaben, die eine *Informationsverarbeitung* erfordern, weisen bereits Abweichungen in der Aufgabensituation auf, gelernte Inhalte können aber sinngemäß wie sie gelernt worden sind angewandt werden. Erst wenn zur Bearbeitung einer Aufgabe *Informationen zu erzeugen* sind, ist der eigenständige kognitive Beitrag hoch, da Lernende das bestehende Wissen nicht unverändert übernehmen können und „selbstständig für spezifische Sachverhalte neue Strukturen“ schaffen müssen (Metzger et al., 1993, S. 81). In der Klassifikationsmatrix für ULME II wurde die Bezeichnung der drei Kategorien zu *Reproduzieren*, *Anwenden/Verstehen* und *Kritisieren/Reflektieren* geändert. Abbildung 11 fasst die Überlegungen von Hofmeister (2005) zusammen.

Wissensart \ Leistungsdimension		Reproduzieren	Anwenden/ Verstehen	Kritisieren/ Reflektieren
Faktenwissen	Terminologisch			
	Empirisch			
Konzeptwissen	Kategorien & Klassifikationen			
	Prinzipien & Verallg.			
Prozedurales Wissen	Fertigkeiten & Verfahren			
	Techniken & Methoden			

Abbildung 11: Klassifikationsmatrix in ULME nach Brand, Hofmeister, & Tramm (2005) sowie Hofmeister (2005)

Es hat sich gezeigt, dass bestehende Kategorisierungsansätze für den Zweck der Untersuchung der vorliegenden Arbeit wenig geeignet zu sein scheinen. Daher soll eine Einordnung der in der Befragung berücksichtigten Aufgaben nach dem Kriterium der zur Lösung der Aufgabe benötigten Wissensart erfolgen. Dabei werden lediglich die vier grundlegenden Wissensarten ohne die von Anderson & Krathwohl (2001) skizzierten Subkategorien herangezogen, um eine zu

feingliedrige und im Rahmen der später vorzunehmenden Selbsteinschätzung der Lernenden vermutlich wenig trennscharfe Unterteilung zu vermeiden.

Die Zuordnung der zur Aufgabenbearbeitung benötigten Wissensarten erscheint für die Textaufgaben zunächst wenig trennscharf zu sein. Die Fragen nach dem Verständnis des Prinzips der getrennten Kontenführung (TA131) und des Belegprinzips (TA133) enthalten Elemente konzeptuellen Wissens (Darstellung von Prinzipien). Die zugrundeliegenden kognitiven Prozesse sind jedoch reproduktiver Natur. Gefragt wird nach der reinen Wiedergabe der genannten Prinzipien, ohne die hierin enthaltenen Zusammenhänge der basalen Elemente (z. B. Aufwand und Ertrag) erläutern zu müssen. Die Frage nach den Tatbestandselementen steuerbarer Umsätze (TA132) beinhaltet die Wiedergabe der in §1 UStG genannten Vorschriften. Ein Zusammenhang basaler Elemente ist hierbei nicht zu erkennen, sodass die Beantwortung dieser Frage Faktenwissen erfordert, das reproduziert werden soll. Alles in allem sind die Textaufgaben überwiegend auf Reproduktion ausgerichtet und mit Faktenwissen zu beantworten.

Die MC-Aufgaben (MC141-MC147) beinhalten die Zuordnung von möglichen Geschäftsvorfällen zu vorgegebenen Buchungssätzen. Um diese Aufgaben adäquat bearbeiten zu können, müssen die Studierenden über hinreichendes konzeptuelles Wissen zum Rechnungswesen verfügen (z. B. Zusammenhänge von Soll und Haben, Aufwendungen und Erträge sowie Aktiva und Passiva). Dieses Wissen wird im Rahmen der MC-Aufgaben nicht nur reproduziert, sondern in dem Sinne auch angewandt, dass die Studierenden basale Wissens Elemente zueinander in Beziehung setzen müssen.

Die Buchungssätze (BS161-BS166) erfordern von den Studierenden prozedurales Wissen zur Bearbeitung. Der zugrundeliegende kognitive Prozess bezieht sich auf das Anwenden von Wissen. Mit der Transferaufgabe (FS181) wenden die Studierenden konzeptuelles Wissen (z. B. Geschäftsjahresgewinnermittlung, Eigenkapitalbestimmung) an.

Es konnte gezeigt werden, dass sich die ausgewählten Aufgaben sowohl hinsichtlich der angesprochenen Wissensart (Faktenwissen, konzeptuelles Wissen, prozedurales Wissen), als auch der zur Bearbeitung benötigten kognitiven Prozesse (Reproduktion vs. Anwendung/Verstehen) unterscheiden. Diese Unterschiede in den Anforderungen müssen sich in den Items zur Selbsteinschätzung wiederfinden lassen, um über einen Vergleich der Selbsteinschätzungen mit der gezeigten Performanz bei der Bearbeitung der Aufgaben Aussagen über die Akkuratheit der Selbsteinschätzungen treffen zu können.

5.2.2.2 Entwicklung von Items zur Erfassung der Selbsteinschätzung

In Kapitel 3.3.1 wurde resümiert, dass die Akkuratheit der Selbsteinschätzungen zu verschiedenen Initiierungszeitpunkten geprüft werden soll. Weiterhin erscheint es sinnvoll, diese auf Aufgaben mit verschiedenen kognitiven Anspruchsniveaus anzupassen und möglichst an konkret messbaren Kriterien auszurichten.

Die Selbsteinschätzungen werden zu beiden Messzeitpunkten sowohl vor (Initiierungszeitpunkt I) als auch nach der Bearbeitung der Aufgaben (Initiierungszeitpunkt II) erhoben. Die Selbsteinschätzungen vor der Aufgabenbearbeitung beziehen sich dabei auf den gesamten Aufgabenkomplex, während die Selbsteinschätzungen nach der Aufgabenbearbeitung sich einerseits unmittelbar auf die soeben bearbeitete Aufgabe (Überzeugungen) und andererseits auf den gesamten bearbeiteten Aufgabenkomplex beziehen (wiederholte Abfrage der Items aus Initiierungszeitpunkt I).

Die Selbsteinschätzung unmittelbar nach der Aufgabenbearbeitung erfolgt immer im gleichen Wortlaut und zielt auf die Überzeugung ab, die vorangegangene Aufgabe korrekt gelöst zu haben. Diese Aussage schätzen die Studierenden auf einer sechsstufigen Skala mit den Polen *1 = gar nicht überzeugt* und *6 = sehr überzeugt* ein. Platziert wird dieses Item nach Bearbeitung einer jeden Textaufgabe sowie nach der Bearbeitung der MC-Aufgaben, der Buchungsaufgaben und der Transferaufgabe. Aus Gründen der Testökonomie wird jedoch darauf verzichtet, die Überzeugung der Richtigkeit jedes aufgestellten Buchungssatzes und jeder einzelnen MC-Aufgabe zu erfassen.

Die Konstruktion der Items zur Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) gestaltet sich komplexer, da diese nicht auf eine spezifische Aufgabe bezogen sind, aber dennoch die zur Bearbeitung der eingesetzten Aufgaben erforderlichen Wissensarten enthalten sollen. Darüber hinaus muss gewährleistet sein, dass die Studierenden den Inhalt der einzuschätzenden Aussagen auch verstehen. Daher werden die Aussagen mit Beispielen versehen, die die unterschiedlichen Niveaustufen für die Studierenden unterscheidbarer machen sollen.

In Tabelle 7 sind die Items zur Selbsteinschätzung vor und nach der Aufgabenbearbeitung (Initiierungszeitpunkt I und II) dargestellt und den entsprechenden Aufgaben zugeordnet. Die erste Aussage (SE122 bzw. SE151) bezieht sich darauf, noch nicht über das nötige Verständnis für die Grundlagen der Buchführung zu verfügen. Diesem Item sind keine Aufgaben unmittelbar zugeordnet. Es wird erwartet, dass eine hohe Ausprägung an dieser Stelle in Zusammenhang mit niedrigen Performanzwerten bei allen Aufgaben steht.

Item ²⁹	Itemformulierung	zugeordnete Aufgaben
SE122 SE151	Mir fehlt noch das Verständnis für die Grundlagen der Buchführung.	-
SE123 SE152	Ich bin in der Lage, grundlegende Begriffe und Prinzipien der Buchführung unabhängig voneinander in wenigen Sätzen wiederzugeben (z. B. Definition erfolgswirksamer Geschäftsvorfälle).	Textaufgaben 1 - 3
SE124 SE153	Ich bin in der Lage, die Zusammenhänge zwischen grundlegenden Begriffen und Prinzipien der Buchführung zu erläutern (z. B. Bestimmung erfolgswirksamer Geschäftsvorfälle).	MC-Aufgaben
SE125 SE154 SE171	Ich bin in der Lage, Buchungssätze zu gegebenen Geschäftsvorfällen aufzustellen.	Buchungsaufgaben
SE126 SE155 SE191	Ich habe grundlegende Begriffe und Prinzipien der Buchführung verstanden und bin in der Lage, sie auf neuartige Aufgabenstellungen anzuwenden.	Transferaufgabe

Tabelle 7: Zuordnung der Items zur Selbsteinschätzung zu den äquivalenten Aufgaben

Mit der zweiten Aussage (SE123 bzw. SE152) wird auf die Wiedergabe von Faktenwissen abgezielt, wie es mehrheitlich in den drei Textaufgaben gefordert wird. Die Anwendung konzeptuellen Wissens wird in den MC-Aufgaben gefordert und mit dem dritten Item (SE124 bzw. SE153) zur Selbsteinschätzung erfasst. Das Aufstellen von Buchungssätzen (prozedurales Wissen) ist Gegenstand der vierten Aussage (SE125 bzw. SE154 und SE171) und bezieht sich entsprechend auf die Aufgaben BS161 – BS167. Der Wissenstransfer auf neuartige Aufgabenstellungen ist mit dem fünften Item zur Selbsteinschätzung abgedeckt (SE126 bzw. SE155 und SE191).

Die skizzierten Items zur Selbsteinschätzung werden in zwei Varianten von den Studierenden erfragt: zunächst im Rahmen eines Single-Choice-Items (SE121) mit allen Taxonomiestufen sowie anschließend über eine sechsstufige Likert-Skala mit den Polen *1 = gar nicht zutreffend* bis *6 = sehr zutreffend*.

²⁹ Angegeben sind die Itembezeichnungen für den ersten (SE122 – SE126) und zweiten Initiierungszeitpunkt (SE151 – SE155, SE171, SE191) der Selbsteinschätzung zum ersten Messzeitpunkt. Die Itembezeichnungen zum zweiten Messzeitpunkt gestalten sich analog hierzu (SE222 bzw. SE251, etc.).

5.2.2.3 Entwicklung von Items zur Erfassung der eingesetzten Lernstrategien

Um erste mögliche Anhaltspunkte zu generieren, wodurch sich eine eventuell vorliegende Fehleinschätzung der Studierenden hinsichtlich ihres Wissens- und Könnensstandes auf die Performanz auswirken könnte, werden von den Studierenden die in den Wochen vor der Datenerhebung eingesetzten Lernstrategien via Selbsteinschätzung durch die Studierenden erfasst. Möglicherweise lassen sich in der Ausprägung der eingesetzten Lernstrategien Unterschiede eruieren, die eine geringere bzw. höhere Performanz erklären könnten. Hierzu werden zunächst bestehende Instrumente zur Erfassung selbstregulierten Lernens dargestellt, um hieraus Implikationen für die Generierung der Items zur Erfassung der eingesetzten Lernstrategien ziehen zu können.

Roth, Ogrin, & Schmitz (2015) geben einen systematischen Überblick über bestehende Instrumente zur Erfassung selbstregulierten Lernens und ihrer Verbreitung in der Forschung innerhalb der vergangenen 30 Jahre. Sie unterscheiden zwischen vier Formen der Datenerhebung, die bei der Erfassung des selbstregulierten Lernens bislang eingesetzt wurden. Der verbreitetste Ansatz ist die schriftliche Befragung der Probanden. Besonders häufig eingesetzt wurde bislang das *Motivated Strategies for Learning Questionnaire* (MSLQ) von Pintrich, Smith, Garcia, & McKeachie (1991) und das *Learning and Study Strategies Inventory* (LASSI) von Weinstein & Palmer (2002). Trotz der angeführten Kritik, dass die auf Likert-Skalen basierenden Instrumente die Probanden dergestalt leiten, dass sie die Nützlichkeit der gegebenen Lernstrategien einschätzen und nicht die tatsächliche Realisierung, werden die weiteren Ansätze zur Erfassung des selbstregulierten Lernens seltener eingesetzt. Für die Erfassung des selbstregulierten Lernens wird alternativ auf Interviews, Techniken des *think-aloud* und dem Einsatz von Lerntagebüchern zurückgegriffen.

Im Folgenden werden die beiden am häufigsten eingesetzten Instrumente zur schriftlichen Befragung kurz dargestellt und anschließend auf ihre Eignung für die Prüfung von Hypothese 3 und 4 analysiert.

Das MSLQ von Pintrich et al. (1991) besteht aus 15 Subskalen zu Motivation (z. B. *Zielorientierung*) und Lernstrategien (z. B. *Zeit und Lernumgebung*) von Studierenden und umfasst insgesamt 81 Items. Diese werden dabei auf einen spezifischen, universitären Kurs bezogen erfragt. Mit dem Instrument wird demnach nicht versucht, kursunabhängig das selbstregulierte Lernen zu erfassen. Die Reliabilität ist nicht für alle Subskalen (z. B. *Extrinsic Goal Orientation*, *Help-Seeking*) zufriedenstellend (Pintrich et al., 1991, S. 10, 29). Tabelle 8 gibt einen

Überblick über die Subskalen zu den Lernstrategien, die für diese Befragung relevant sein könnten.

Vor dem Hintergrund der eingesetzten Aufgaben (s. Kapitel 5.2.2.1) sind die Strategien des *Rehearsal* insbesondere für die Aufgaben, die eine Reproduktion von den Lernenden erfordern, von Bedeutung. *Elaborationsstrategien* kommen vermehrt zum Tragen, wenn Anwendungs- und Transferaufgaben gelöst werden müssen. Strategien zur *Organisation* sind zu allgemein und unspezifisch, dass eine unterschiedliche Ausprägung in Abhängigkeit der Selbsteinschätzung erwartet werden kann. Die Subskala zum *Critical Thinking* wird nicht weiter berücksichtigt, da die hier abgebildeten Sachverhalte nur wenige Gemeinsamkeiten mit den in der Veranstaltung geforderten aufweisen. Die Items zur *metakognitiven Selbstregulation* bilden die zentralen Prozesse (Planung, Monitoring, Kontrolle) selbstregulierten Lernens ab, die aber nicht zwingend im Zusammenhang mit der Akkuratheit der Selbsteinschätzung stehen. Die Subskalen zu *Zeit und Lernumgebung* sowie der *Regulierung der Anstrengung* sind wiederum zu unspezifisch und allgemein, dass sie nicht weiter berücksichtigt werden. Dagegen sind die Items über das *Lernen mit Kommilitonen* und das *Help-Seeking* grundsätzlich geeignet, Unterschiede zwischen Studierenden mit akkurater und inakkurater Selbsteinschätzung zu eruieren. Somit bieten die Subskalen *Rehearsal*, *Elaborationsstrategien*, *Lernen mit Kommilitonen* und *Help-Seeking* Anknüpfungspunkte für die Erfassung von Lernstrategien, deren Ausprägung von der Akkuratheit der Selbsteinschätzung abhängig sein könnte.

Subskala	Beschreibung	Beispielitem
Cognitive and Meta-cognitive Strategies: Rehearsal (4 Items)	Strategien des Wiederholens, insbesondere relevant bei Aufgaben, die eine Informationserinnerung erfordern	<i>When I study for this class, I practice saying the material to myself over and over.</i>
Cognitive and Meta-cognitive Strategies: Elaboration (6 Items)	Elaborationsstrategien, die der Speicherung von Informationen im Langzeitgedächtnis dienen	<i>When I study for this class, I pull together information from different sources, such as lectures, readings, and discussions.</i>
Cognitive and Meta-cognitive Strategies: Organization (4 Items)	Organisationsstrategien die dem Lernen helfen, wichtige Informationen auszuwählen und hervorzuheben	<i>When I study the readings for this course, I outline the material to help me organize my thoughts.</i>
Cognitive and Meta-cognitive Strategies: Critical Thinking (5 Items)	Strategien umschreiben den Grad der Anwendung des Gelernten und die kritische Auseinandersetzung mit diesem in neuen Situationen	<i>I often find myself questioning things I hear or read in this course to decide if I find them convincing.</i>
Cognitive and Meta-cognitive Strategies: Metacognitive Self-Regulation (12 Items)	Erfassung der grundlegenden selbstregulativen Tätigkeiten Planung, Überwachung und Regulierung	<i>During class time I often miss important points because I'm thinking of other things.</i>
Resource Management Strategies: Time and Study Environment (8 Items)	Regulierung der Zeit und der Lernumgebung	<i>I usually study in a place where I can concentrate on my course work.</i>
Resource Management Strategies: Effort Regulation (4 Items)	Fähigkeit, Anstrengung zu regulieren und die Aufmerksamkeit auch bei uninteressanten Aufgaben aufrecht zu erhalten	<i>I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do.</i>
Resource Management: Peer Learning (3 Items)	Zusammenarbeit mit Kommilitonen	<i>When studying for this course, I often try to explain the material to a classmate or a friend.</i>
Resource Management: Help Seeking (4 Items)	Hilfersuchen der Studierenden bei Notwendigkeit	<i>I ask the instructor to clarify concepts I don't understand well.</i>

Tabelle 8: Übersicht der Subskalen zu den Lernstrategien nach Pintrich et al. (1991)

Das LASSI wurde entwickelt, um das Bewusstsein der Lernenden über Lernstrategien und deren Nutzung zu untersuchen. Dabei wird zwischen zehn Subskalen der Fähigkeit (z. B. Teststrategien), des Willens (z. B. Einstellung & Motivation) und der Selbstregulation (z. B. Zeitmanagement) unterschieden (s. Tabelle 9). Insgesamt umfasst das Instrument 80 Items. Die Reliabilität ist für alle Subskalen zufriedenstellend. Die Subskalen werden erneut dahingehend geprüft, inwiefern sie Aussagen über die Quantität des Lernens erlauben und von der Akkuratheit der Selbsteinschätzung abhängig zu sein scheinen.

Für die Subskala der *Ängstlichkeit* (anxiety) ist zu vermuten, dass sie von der Akkuratheit der Selbsteinschätzung abhängig sein kann, allerdings enthält sie keine konkreten Lernstrategien. Gleiches gilt für die Subskalen *Einstellung* (attitude), *Konzentration* (concentration), *Organisation* (scheduling) und der *Umgang mit Prüfungen* (test strategies). Die verbliebenen Subskalen scheinen für einen weiteren Einsatz geeignet zu sein. Diese sind vergleichbar mit der von Pintrich et al. (1991) entwickelten Subskala zur metakognitiven Selbstregulation.

Subskala	Beschreibung	Beispielitem
Anxiety	Sorgen bezüglich Schule/Universität	<i>Worrying about doing poorly interferes with my concentration on tests.</i>
Attitude	Einstellung zu und Interesse für Schule/Universität	<i>Success in school is very important to me.</i>
Concentration	Fähigkeit zu konzentriertem Arbeiten	<i>I often find that I have been reading, but don't know what it was about.</i>
Information Processing	Einsatz von Elaborationsstrategien	<i>I try to find relationships between what I am learning and what I already know.</i>
Motivation	Bereitschaft zu Lernen Motivation	<i>I read the textbook assigned for my classes.</i>
Scheduling	Organisation	<i>I only study when there is the pressure of a test.</i>
Selecting the main idea	Konzentration auf wesentliche Aspekte des zu Lernenden, Auswahl relevanter Inhalte	<i>My underlining is helpful when I review text material.</i>
Self-Testing	Reflexion der gelernten Inhalte	<i>I stop periodically while reading and mentally go over or review what was said.</i>
Study Aids	Einsatz von Lernstrategien	<i>When they are available, I attend group review sessions.</i>
Test Strategies	Umgang mit Prüfungen	<i>I think through the meaning of test questions before I begin to answer them.</i>

Tabelle 9: Übersicht der Subskalen des LASSI (Weinstein, Zimmermann, & Palmer, 1988)

Die Darstellung der beiden Instrumente zur Erfassung selbstregulierten Lernens offenbarte, dass nicht alle Subskalen für den Zweck der Prüfung der Zusammenhänge zwischen akkurater Selbsteinschätzung und Lernstrategieeinsatz geeignet zu sein scheinen. Nur wenige Items beziehen sich auf konkrete Lernaktivitäten, die vorrangig erfasst werden sollen. Schließlich wird vermutet, dass es abhängig von der Selbsteinschätzung ist, wie intensiv bestimmte Lernstrategien eingesetzt werden. Da überdies mit dem wiederholten Einsatz der Items zur Selbsteinschätzung und der Bearbeitung von Aufgaben des externen Rechnungswesens der zu bearbeitende Fragebogen bereits komplex ist, wurde davon Abstand genommen, ein Instrument vollständig einzusetzen.

Es schien zweckmäßig, eigene Items in Anlehnung an die zuvor beschriebenen Subskalen zu entwickeln. In Tabelle 10 sind die Items und die vorgenommene Kategorisierung dargestellt. Es ist beabsichtigt, mit 18 Items vier Kategorien abzubilden:

- i) Veranstaltungsbesuch (*3 Items*)
- ii) Vor- und Nachbereitung der Inhalte (*6 Items*)
- iii) Lösen von Übungsaufgaben (*5 Items*)
- iv) Umgang mit Schwierigkeiten (*4 Items*).

In einem ersten Schritt wird der Veranstaltungsbesuch in Vorlesung, Übung und Tutorium erfragt. Schließlich erscheint es denkbar, dass bspw. eine hohe Überschätzung des eigenen Wissens- und Könnensstandes dazu führt, dass Veranstaltungen nicht regelmäßig besucht werden. Darüber hinaus soll allgemein eruiert werden, inwiefern die Studierenden der Ansicht sind, dass sie die Veranstaltungsinhalte aus Vorlesung, Übung und Tutorium vor- und nachbereiten. Das *Lösen von Übungsaufgaben* knüpft an der Subskala *Self-Testing* des *LASSI* an, wohingegen der *Umgang mit Schwierigkeiten* mit den Skalen des *Peer Learning* und *Help Seeking* aus dem *MSLQ* Parallelen aufweist.

Kategorie	Item	Itemformulierung
Veranstaltungs- besuch	LS1101	<i>Ich habe die Vorlesung regelmäßig besucht.</i>
	LS1102	<i>Ich habe die Übung regelmäßig besucht.</i>
	LS1103	<i>Ich habe die Tutorien regelmäßig besucht.</i>
Vor- und Nach- bereitung der Inhalte	LS1104	<i>Ich bereite die Inhalte der Vorlesung vor.</i>
	LS1105	<i>Ich bereite die Inhalte der Übung vor.</i>
	LS1106	<i>Ich bereite die Inhalte der Tutorien vor.</i>
	LS1107	<i>Ich bereite die Inhalte der Vorlesung nach.</i>
	LS1108	<i>Ich bereite die Inhalte der Übung nach.</i>
	LS1109	<i>Ich bereite die Inhalte der Tutorien nach.</i>
Lösen von Übungsaufga- ben	LS11010	<i>Ich versuche, die bereitgestellten Übungsaufgaben eigenständig zu lösen.</i>
	LS11011	<i>Ich bearbeite die bereitgestellten Übungsaufgaben gemeinsam mit Kommilitonen.</i>
	LS11012	<i>Bevor ich Übungsaufgaben löse, gehe ich die Vorlesungsunterlagen nochmals durch.</i>
	LS11013	<i>Wenn ich einfachere Übungsaufgaben sicher lösen kann, wende ich mich rasch schwierigeren Aufgaben zu.</i>
	LS11014	<i>Ich wiederhole oft einfache Übungsaufgaben, auch wenn ich sie schon mehrmals richtig gelöst habe.</i>
Umgang mit Schwierigkeiten	LS11015	<i>Wenn ich bei einer Aufgabe nicht weiterkomme, stelle ich diese zunächst zurück, um sie später zu lösen.</i>
	LS11016	<i>Wenn ich bei einer Aufgabe nicht weiterkomme, frage ich Kommilitonen um Rat.</i>
	LS11017	<i>Wenn ich bei einer Aufgabe nicht weiterkomme, gehe ich die vorlesungsbegleitende Literatur noch einmal genau durch.</i>
	LS11018	<i>Wenn ich bei einer schwierigen Übungsaufgabe nicht weiterkomme, bearbeite ich zunächst einfachere Aufgaben.</i>

Tabelle 10: Übersicht der Items zum Einsatz von Lernstrategien

5.2.2.4 Entwicklung von Items zur Erfassung des Zeitaufwands

Vor dem Hintergrund des noch nicht abschließend geklärten Einflusses der Lernzeit auf die Fähigkeit zur Selbsteinschätzung (vgl. Andrade et al., 2010; Dunlosky & Rawson, 2012; Kostons et al., 2010) soll zum ersten Initiierungszeitpunkt der für das Lernen beabsichtigte Zeitaufwand im Rahmen der Veranstaltung erfasst werden. Um möglichst genaue Angaben von den Studierenden gewinnen zu können, wird der Zeitaufwand auf zwei verschiedene Art und Weisen erfasst. Zum einen werden die Studierenden gebeten, diesen auf einer fünfstufigen Likertskala mit den Polen *1 = gar nicht zutreffend* und *5 = sehr zutreffend* im Vergleich zu anderen Lehrveranstaltungen ihres Studiums zu bewerten (ZA1111). Zum anderen werden die aufgewandten Zeiten für die Vor- und Nachbereitung von Vorlesung, Übung und Tutorium erfasst (ZA1112 – ZA1114). In Tabelle 11 sind die Items zur Erfassung des Zeitaufwands abgebildet.

Item	Itemformulierung
ZA1111	Ich habe im Rahmen der Veranstaltung "Grundlagen des externen Rechnungswesens" in den vergangenen Wochen im Vergleich zu anderen Veranstaltungen meines Studiums sehr viel mehr Zeit investiert.
ZA1112	Zeit für Vor- und Nachbereitung der Vorlesung insgesamt [Angabe in Minuten pro Woche]
ZA1113	Zeit für Vor- und Nachbereitung der Übung insgesamt [Angabe in Minuten pro Woche]
ZA1114	Zeit für Vor- und Nachbereitung der Tutorien insgesamt [Angabe in Minuten pro Woche]

Tabelle 11: Übersicht der Items zur Erfassung des Zeitaufwands

5.2.2.5 Allgemeine Angaben

Die allgemeinen Angaben umfassen Geschlecht, Alter, Studienfach, Typ der erworbenen Hochschulreife sowie die Vorbildung im Bereich des externen Rechnungswesens. Außerdem wurde im Rahmen eines Items die Motivation zum Lernen von Inhalten des Rechnungswesens erfragt.

Alle Items des Instrumentes sowie ihre Reihung sind dem Codebuch in Anhang A2 zu entnehmen.

5.2.3 Ablauf der Datenerhebung

Die Daten wurden im Rahmen der regulären 90-minütigen Übung zur Veranstaltung „Grundlagen des externen Rechnungswesens“ erhoben. Angeboten werden in jeder Woche zwei inhaltsgleiche Übungen, sodass die Studierenden nur eine Übung in der Woche besuchen. Die Befragung fand in den ersten 45 Minuten der Übung zu beiden Übungsterminen statt.

Zu Beginn der Befragung wurden die Studierenden über den Gegenstand der Untersuchung instruiert. Die Instruktionen beinhalteten die Vorstellung des Testleiters sowie die Klärung der Rahmenbedingungen des Testablaufs. Diese umfassten die Bekanntgabe der Bearbeitungsdauer von 45 Minuten, die Aufforderung, keine Hilfsmittel zu verwenden, die Zusicherung der Anonymität der Daten sowie der Erklärung zur Bildung des abgefragten Codes. Der Code setzt sich zusammen aus der Summe der ersten drei Ziffern der Postleitzahl des Heimatortes vor Studienbeginn, der Summe aus Geburtstag und Geburtsmonat der Mutter und der Summe der letzten drei Ziffern der Matrikelnummer. Auf Basis der gemachten Angaben konnten keine Rückschlüsse auf die Identität der Studierenden gezogen werden. Die Abfrage eines persönlichen Codes war notwendig, um die Angaben der Studierenden über beide Erhebungszeitpunkte vergleichen zu können.

Anschließend wurden die Fragebögen vom Testleiter, zwei wissenschaftlichen Mitarbeitern und zwei wissenschaftlichen Hilfskräften an die Studierenden ausgegeben. Die Bearbeitungszeit begann, nachdem alle Studierenden einen Fragebogen erhielten. Das Bearbeitungsende wurde für alle Studierenden verbindlich nach 45 Minuten gesetzt, eine frühere Abgabe des Fragebogens war nicht möglich. Abschließend wurden die Fragebogen eingesammelt und mit der regulären Übung fortgefahren.

5.3 Designkomponenten des zweiten Messzeitpunktes

5.3.1 Stichprobe

Zum zweiten Erhebungstermin in der letzten Vorlesungswoche des Semesters nahmen an beiden Übungsterminen insgesamt 163 Studierende teil (s. Tabelle 12). Zum ersten Übungstermin konnte ein Rücklauf von 122 Fragebögen registriert werden, zum zweiten Übungstermin konnten noch 41 Studierende für die Teilnahme an der Befragung gewonnen werden. Die Studierenden waren im Durchschnitt 21.26 Jahre alt ($SD = 1.87$) und mehrheitlich weiblich ($n = 73$ Studierende³⁰). Dabei sind sie in der Mehrzahl in den Bachelorstudiengängen „Kultur und Wirtschaft“ ($n = 46$), „Unternehmensjurist“ ($n = 23$) und „Wirtschaftspädagogik“ ($n = 19$) eingeschrieben. Die Hochschulreife wurde überwiegend am allgemeinbildenden Gymnasium erworben ($n = 86$). Über Vorkenntnisse im Rechnungswesen dürften nur wenige Studierende verfügen. Die Probanden hatten in wenigen Fällen bereits eine kaufmännische Ausbildung absolviert ($n = 10$) und kaum schulische Erfahrungen mit dem Rechnungswesen aufzuweisen ($n = 17$).

³⁰ Zum zweiten Messzeitpunkt ist der Anteil an fehlenden demographischen Angaben hoch, da sie von jenen Studierenden, die bereits am ersten Erhebungstermin teilgenommen haben, nicht erneut erfragt wurden. Aus diesem Grund wird in der Folge die Häufigkeit der Nennungen für die weiteren demographischen Angaben anstelle von relativen Werten angegeben.

Kriterium		Übungs- termin I	Übungs- termin II	Σ
N		122	41	163
Geschlecht	weiblich	60	13	73
	männlich	27	13	40
	fehlende Angabe	35	15	50
Alter	M	21.20	21.42	21.26
	SD	1.88	1.86	1.87
	fehlende Angabe	43	15	58
Studienfach	Bachelor Kultur und Wirtschaft	34	12	46
	Bachelor Unternehmensjurist	20	3	23
	Bachelor VWL	7	3	10
	Bachelor Wirtschaftsinformatik	0	3	3
	Bachelor Wirtschaftsmathematik	4	0	4
	Bachelor Wirtschaftspädagogik	17	2	19
	Lehramt an Gymnasien	3	0	3
	Master Psychologie	0	0	0
	sonstiges	3	2	5
	fehlende Angabe	34	16	50
HZB	allgemeinbildendes Gymnasium	69	17	86
	Wirtschaftsgymnasium	8	4	12
	anderer Schultyp	7	2	9
	fehlende Angabe	38	18	56
abgeschlossene kaufmännische Berufsausbildung	ja	8	2	10
	nein	79	22	101
	fehlende Angabe	35	17	52
schulische Erfahrun- gen mit Rechnungs- wesen	ja	14	3	17
	nein	73	21	94
	fehlende Angabe	35	17	52

Tabelle 12: Charakterisierung der Stichprobe zu Messzeitpunkt II

In Tabelle 13 ist die Stichprobe, die an beiden Messzeitpunkten teilgenommen hat (Teilstichprobe), dargestellt. Insgesamt konnte anhand des Codes eine eindeutige Zuordnung für 56 Studierende vorgenommen werden. Die Mehrheit der Probanden war demnach weiblich ($n = 42$ Studierende) und im Durchschnitt $M = 20.76$ Jahre alt ($SD = 1.65$). Die meisten Studierenden, die an beiden Messzeitpunkten an der Befragung teilnahmen, befanden sich im Bachelorstudiengang „Kultur und Wirtschaft“ ($n = 26$ Studierende) und erlangten die HZB an einem allgemeinbildenden Gymnasium ($n = 48$ Studierende). Dementsprechend niedrig ist die Zahl der Studierenden mit einer abgeschlossenen kaufmännischen Berufsausbildung ($n = 5$ Studierende) respektive mit schulischen Erfahrungen im Rechnungswesen ($n = 9$ Studierende).

Kriterium		Σ
N		56
Geschlecht	weiblich	42
	männlich	14
Alter	M	20.76
	SD	1.65
	fehlende Angabe	2
Studienfach	Bachelor Kultur und Wirtschaft	26
	Bachelor Unternehmensjurist	8
	Bachelor VWL	3
	Bachelor Wirtschaftsinformatik	0
	Bachelor Wirtschaftsmathematik	0
	Bachelor Wirtschaftspädagogik	10
	Lehramt an Gymnasien	1
	Master Psychologie	0
	sonstiges	8
HZB	allgemeinbildendes Gymnasium	48
	Wirtschaftsgymnasium	5
	anderer Schultyp	3
abgeschlossene kaufmännische Berufsausbildung	ja	5
	nein	51
schulische Erfahrungen mit Rechnungswesen	ja	9
	nein	47

Tabelle 13: Charakterisierung der Teilstichprobe

5.3.2 Instrument

Um die Selbsteinschätzungen und die Performanz über die beiden Erhebungszeitpunkte vergleichen zu können, wurde der Fragebogen zum zweiten Erhebungszeitpunkt nicht geändert. Allerdings mussten Studierende, die am ersten Erhebungszeitpunkt teilnahmen, die allgemeinen Angaben nicht erneut beantworten.

5.3.3 Ablauf der Datenerhebung

Der grundsätzliche Ablauf der Datenerhebung unterscheidet sich von jenem des ersten Erhebungszeitpunktes nicht. Allerdings wurden die Rahmenbedingungen in zwei Punkten angepasst: Da sich die Bearbeitungsdauer von 45 Minuten am ersten Erhebungszeitpunkt als zu lang herausgestellt hat, wurde diese auf 30 Minuten reduziert. Der Bearbeitungsgrad der Fragebögen zeigt, dass diese Zeitvorgabe für eine vollständige Bearbeitung des Fragebogens ebenfalls ausreichte. Außerdem wurde am zweiten Erhebungszeitpunkt ein monetärer Anreiz gesetzt, um die Studierenden zur Teilnahme zu motivieren. So wurden die vier besten Aufgabenbearbeitungen mit jeweils 50 € ausgezeichnet.

5.4 Datenauswertung

Die Fragebögen wurden für beide Messzeitpunkte mit EpiData³¹ erfasst. Die Eingabe erfolgte auf der Grundlage des Codebuchs, das im Anhang A2 dargestellt ist. Um die Performanz feststellen zu können, wurden die Antworten der Studierenden bezüglich der zu bearbeitenden Aufgabenstellungen korrigiert und bepunktet. Hierfür wurde eine Musterlösung erstellt, die dem Codebuch zu entnehmen ist. Insgesamt konnten die Studierenden bei den Testaufgaben 34 Punkte erreichen. Hiervon entfallen elf Punkte auf die Textaufgaben, sieben Punkte auf die MC-Aufgaben, zwölf Punkte auf die Buchungsaufgaben und vier Punkte auf die Transferaufgabe.

Die Auswertung der gewonnenen Daten erfolgt in mehreren Schritten. Um einen Überblick über die Ausprägung der studentischen Selbsteinschätzungen und der Performanz zu erhalten, werden für alle Initiierungszeitpunkte der Selbsteinschätzung beider Messzeitpunkte Mittelwerte (M), Standardabweichungen (SD) und Häufigkeitsverteilungen bestimmt. Darüber hinaus werden die Veränderungen in den Selbsteinschätzungen und der Performanz zwischen den Initiierungszeitpunkten mittels t-Tests für verbundene Stichproben auf Signifikanz geprüft.

Danach wird die Akkuratheit der Selbsteinschätzung an den verschiedenen Initiierungs- und Messzeitpunkten ermittelt. Als Erstes erfolgt hierfür die Recodierung der Selbsteinschätzungen von *1 = gar nicht zutreffend* bis *6 = sehr zutreffend* auf *0 = gar nicht zutreffend* bis *5 = sehr zutreffend* und die Normierung der erreichten Punktzahlen auf einen maximalen Wert von fünf. Anschließend werden drei Akkuratheitsmaße der Selbsteinschätzung bestimmt. Durch Subtraktion der Performanzwerte von den Werten der Selbsteinschätzungen kann die absolute Abweichung zwischen Selbsteinschätzung und tatsächlich realisierter Performanz als erstes Maß für die Akkuratheit der Selbsteinschätzung berechnet werden. Ein Wert von oder nahe Null ließe auf eine akkurate Selbsteinschätzung, negative Werte auf Unterschätzungs- sowie positive Werte auf Überschätzungstendenzen schließen. Im weiteren Verlauf werden die so ermittelten Abweichungen auch betragsmäßig erfasst (zweites Akkuratheitsmaß), um eine allgemeine Unterscheidung zwischen akkuraterer und inakkuraterer Selbsteinschätzung vornehmen zu können. Als drittes Akkuratheitsmaß für die Selbsteinschätzungen werden die Korrelationen nach Pearson zwischen Selbsteinschätzung und tatsächlicher Performanz berechnet. Die Berechnun-

³¹ EpiData ist eine Freeware-Software aus Dänemark, die zur Datenerfassung eingesetzt werden kann und über verschiedene Kontrollabfragen (z. B. double entry verification) Eingabefehler minimieren soll. Weitere Informationen finden sich unter <http://www.epidata.dk/>.

gen werden sowohl für die Gesamtstichprobe als auch lediglich für jene Studierende vorgenommen, die an beiden Messzeitpunkten an der Befragung teilnahmen. In Tabelle 14 ist das Vorgehen bei der Datenauswertung zusammengefasst.

Ziel	statistische Prozeduren/Vorgehen	MZ/IZP
Ausprägung von Selbsteinschätzung und Performanz (Kapitel 6.1)	<ul style="list-style-type: none"> - Häufigkeitsverteilungen - Mittelwerte/Standardabweichungen - t-Tests für verbundene Stichproben zur Überprüfung der Veränderungen zwischen Initiierungs- und Messzeitpunkten der Selbsteinschätzungen sowie der Performanz 	<ul style="list-style-type: none"> - Für alle Mess- und Initiierungszeitpunkte der Selbsteinschätzung
Akkuratheit der Selbsteinschätzung (Kapitel 6.2)	<ul style="list-style-type: none"> - Bestimmung der Performanzwerte in Abhängigkeit der Einstufung des Wissens und Könnens durch die Studierenden zu Beginn der Befragungen (SE121 bzw. SE221), Signifikanzprüfung mit Kruskal-Wallis-Test - Normierung der Selbsteinschätzungen und erreichten Punktzahlen <p>Bestimmung der Akkuratheit der Selbsteinschätzung anhand dreier Maße:</p> <ol style="list-style-type: none"> 1) Bestimmung der absoluten Abweichungen (Selbsteinschätzungen - Performanz) 2) Bestimmung der betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz (Berechnung von Mittelwerten und Standardabweichungen) 3) Bestimmung der Korrelation nach Pearson zwischen Selbsteinschätzung und Performanz <ul style="list-style-type: none"> - Prüfung des Einflusses von Geschlecht, Hochschulzugangsberechtigung und Motivation auf die Akkuratheit der Selbsteinschätzung (Berechnung von t-Tests für unabhängige Stichproben) 	<ul style="list-style-type: none"> - Für alle Mess- und Initiierungszeitpunkte der Selbsteinschätzung
Akkuratheit der Selbsteinschätzung ↔ Performanz (Kapitel 6.3)	<ul style="list-style-type: none"> - Bestimmung der Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt (analoges Vorgehen wie in Kapitel 6.2) - Prüfung der Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt und Performanz zum zweiten Messzeitpunkt durch t-Tests - Bildung von Kreuztabellen 	<ul style="list-style-type: none"> - Für Studierende, die an beiden Messzeitpunkten teilnahmen - Für alle Initiierungszeitpunkte der Selbsteinschätzung
Akkuratheit der Selbsteinschätzung ↔ Lernstrategie & Lernzeit (Kapitel 6.4)	<ul style="list-style-type: none"> - Prüfung der Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt und realisierten Lernstrategien sowie der Lernzeit durch t-Tests 	<ul style="list-style-type: none"> - Für alle Mess- und Initiierungszeitpunkte der Selbsteinschätzung

Tabelle 14: Auswertungsplan

6. Ergebnisse

6.1 Ausprägung der Selbsteinschätzungen und Performanz

6.1.1 Selbsteinschätzungen und Performanz der Gesamtstichprobe

Im Folgenden werden unter Berücksichtigung der Gesamtstichprobe zunächst überwiegend deskriptive Ergebnisse (Mittelwerte und Standardabweichungen) zu den Selbsteinschätzungen und der gezeigten Performanz bei der Aufgabenbearbeitung für beide Messzeitpunkte dargestellt, um einen Überblick über die Verteilung der Selbsteinschätzungen und der Ausprägung der Performanz zu erhalten. Anschließend werden verschiedene Maße zur Bestimmung der Akkuratheit der Selbsteinschätzung berechnet (s. Kapitel 6.2). Des Weiteren werden Zusammenhänge zwischen akkurater Selbsteinschätzung und Performanz (s. Kapitel 6.3) und Unterschiede im Einsatz der Lernstrategien sowie der aufgewandten Lernzeit in Abhängigkeit von der Akkuratheit der Selbsteinschätzung analysiert (s. Kapitel 6.4). Die Ergebnisbetrachtung und Hypothesenprüfung erfolgt dabei, wenn möglich, sowohl für die Teilstichprobe, die an beiden Messzeitpunkten an der Befragung teilnahm als auch für die Gesamtstichprobe, die auch Studierende enthält, die nur an einem der beiden Messzeitpunkte anwesend waren.

Selbsteinschätzung

Die Studierenden wurden zu verschiedenen Initiierungszeitpunkten (vor bzw. nach der Aufgabenbearbeitung) in verschiedenen Formen (kriterienorientiert bzw. überzeugungsbezogen) um eine Selbsteinschätzung zu ihrem Wissens- und Könnensstand im Bereich des externen Rechnungswesens gebeten.

Die erste Selbsteinschätzung erfolgte vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) anhand von an Wissensarten ausgerichteten Niveaustufen (SE121³²). Dabei wurden die Studierenden angehalten, die Niveaustufe auszuwählen, die ihren Wissens- und Könnensstand am treffendsten umschreibt. Knapp ein Viertel der befragten Studierenden gab zum ersten Messzeitpunkt an, noch nicht über das nötige Verständnis für die Grundlagen der Buchführung zu verfügen. Der Anteil der Studierenden, die schätzten, grundlegende Begriffe und Prinzipien wiedergeben zu können, liegt bei 15.7 %. 39.1 % der Studierenden sahen sich jedoch bereits in der Lage, Buchungssätze korrekt aufstellen zu können.

³² Der genaue Wortlaut aller Items ist dem Anhang A2 zu entnehmen.

Zum zweiten Messzeitpunkt haben sich die Selbsteinschätzungen (SE221) zugunsten der höheren Niveaustufen verschoben. Nur eine Minderheit gab am Ende des Semesters an, noch nicht über das nötige Verständnis zu verfügen (2.5 % der Studierenden) bzw. grundlegende Begriffe und Prinzipien nur wiedergeben zu können (9.2 % der Studierenden). Mehr als die Hälfte der Studierenden schätzte sich dergestalt ein, Buchungsaufgaben oder Transferaufgaben korrekt bearbeiten zu können. In Abbildung 12 sind die relativen Häufigkeiten der Nennungen für beide Messzeitpunkte zusammenfassend dargestellt.

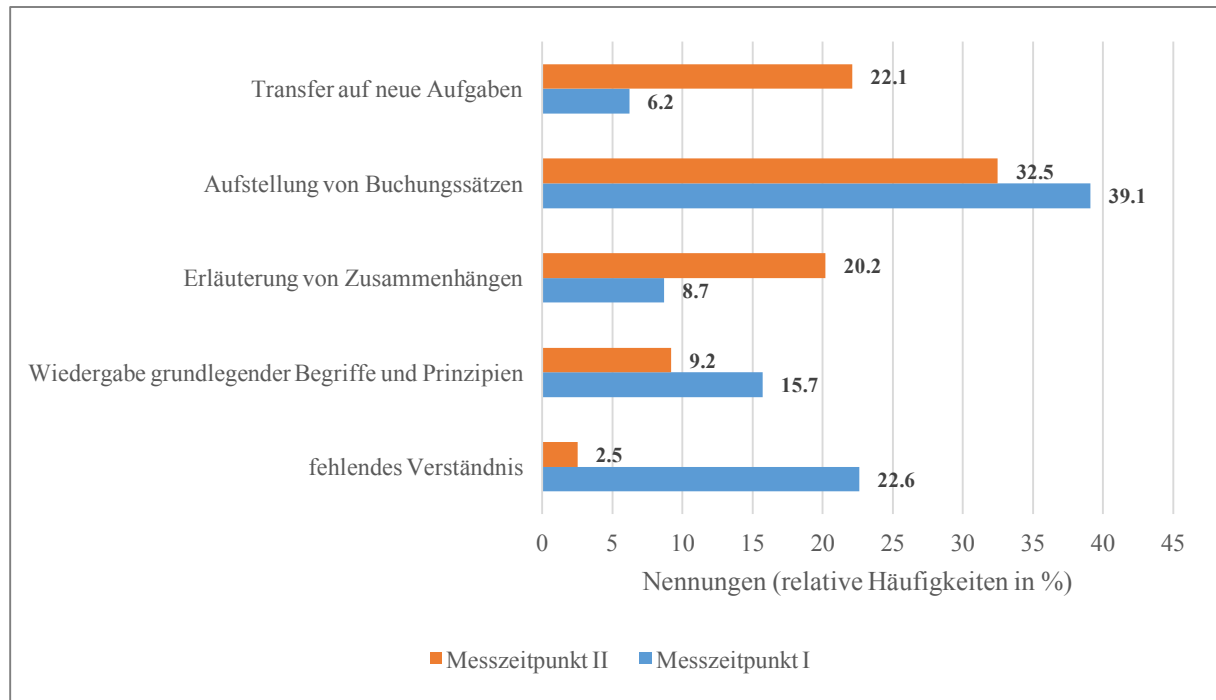


Abbildung 12: Vergleich der Selbsteinschätzungen vor der Aufgabenbearbeitung (SE121 und SE221) an Messzeitpunkt I ($N = 359$) und Messzeitpunkt II ($N = 141$)

Anschließend wurden die einzelnen Niveaustufen nochmals separat über eine sechsstufige Likert-Skala von 1 = *gar nicht zutreffend* bis 6 = *sehr zutreffend* durch die Studierenden eingeschätzt (SE122 – SE126 bzw. SE222 – SE226). Hierbei bestätigen sich die Ergebnisse aus der Betrachtung der Selbsteinschätzung der Items SE121 bzw. SE221. Die Studierenden stimmten der Einschätzung, Buchungssätze aufstellen zu können (SE125 bzw. SE225) zu beiden Messzeitpunkten am stärksten zu. Für alle Items, mit Ausnahme des negativ gepolten Items, noch nicht über das nötige Verständnis zu verfügen (SE122 bzw. SE222), waren die Zustimmungswerte zum zweiten Messzeitpunkt höher als zum ersten Messzeitpunkt (s. Tabelle 15).

	Messzeitpunkt I ($240 \leq n^{33} \leq 382$)		Messzeitpunkt II ($92 \leq n \leq 163$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Initiierungszeitpunkt I³⁴				
SE122: fehlendes Verständnis	2.96	1.55	2.09	1.07
SE123: Wiedergabe grundlegender Begriffe & Prinzipien	3.50	1.33	4.08	1.13
SE124: Erläuterung von Zusammenhängen	3.22	1.31	4.04	1.07
SE125: Aufstellung von Buchungssätzen	3.96	1.53	4.70	1.08
SE126: Transfer auf neue Aufgaben	2.79	1.34	3.59	1.22
Initiierungszeitpunkt II				
SE151: fehlendes Verständnis	3.56	1.60	2.52	1.28
SE152: Wiedergabe grundlegender Begriffe & Prinzipien	2.79	1.24	3.50	1.13
SE153: Erläuterung von Zusammenhängen	2.72	1.25	3.59	1.18
SE154: Aufstellung von Buchungssätzen I	3.23	1.50	4.36	1.10
SE155: Transfer auf neue Aufgaben I	2.43	1.29	3.47	1.25
SE171: Aufstellung von Buchungssätzen II	2.90	1.46	4.63	1.11
SE191: Transfer auf neue Aufgaben II	2.16	1.21	3.34	1.29
Überzeugungen				
SE131: Textaufgabe I	2.41	1.37	3.12	1.56
SE132: Textaufgabe II	1.56	1.15	2.50	1.64
SE133: Textaufgabe III	3.72	1.47	3.99	1.26
SE148: MC-Aufgaben	2.64	1.35	3.96	1.23
SE167: Buchungsaufgaben	2.91	1.54	4.94	1.15
SE182: Transferaufgabe	2.04	1.47	3.01	1.78

Tabelle 15: Mittelwerte und Standardabweichungen der Selbsteinschätzungen

³³ Die hohe Zahl fehlender Werte an beiden Messzeitpunkten resultiert aus nicht angegebenen Überzeugungen, falls die zugehörige Aufgabe zuvor nicht gelöst wurde.

³⁴ Die Itembezeichnungen beziehen sich in dieser Tabelle auf den ersten Messzeitpunkt. Für den zweiten Messzeitpunkt bilden sich diese analog (z. B. SE222, SE223, etc.)

Die gleichen Items zur Selbsteinschätzung wurden nach der Bearbeitung von Text- und MC-Aufgaben³⁵ (Initiierungszeitpunkt II) wiederholt erfragt (SE151 – SE155 bzw. SE251 – SE255). Es zeigt sich wiederum, dass die Selbsteinschätzungen zum zweiten Messzeitpunkt höher sind als zum ersten Messzeitpunkt. Außerdem stimmten die Studierenden erneut der Aussage am stärksten zu, bereits Buchungsaufgaben bearbeiten zu können. Dabei ist die Einschätzung, Buchungsaufgaben lösen zu können zum zweiten Messzeitpunkt für den ersten Initiierungszeitpunkt am höchsten ($M = 4.70$, $SD = 1.08$). Sie sinkt daraufhin zunächst nach der Bearbeitung der Text- und MC-Aufgaben ($M = 4.36$, $SD = 1.10$), um schließlich nach der Bearbeitung der Buchungsaufgaben erneut anzusteigen ($M = 4.63$, $SD = 1.11$). Dagegen sinkt die Einschätzung, Transferaufgaben adäquat lösen zu können kontinuierlich von $M = 3.59$ ($SD = 1.22$) auf $M = 3.34$ ($SD = 1.29$).

Über einen t-Test für verbundene Stichproben³⁶ wurde außerdem geprüft, inwiefern sich die Mittelwerte zu den Selbsteinschätzungen für die verschiedenen Initiierungszeitpunkte an beiden Messzeitpunkten voneinander unterscheiden (s. Anhang A3). Die Selbsteinschätzungen nach der Aufgabebearbeitung veränderten sich am ersten Messzeitpunkt dahingehend, dass die befragten Studierenden der Aussage, noch nicht über das nötige Verständnis der Buchführung zu verfügen, signifikant stärker zustimmten ($M = 3.56$, $t(374) = -9.25$, $p < 0.001$), während sie ihr Wissen und Können bzgl. der weiteren Niveaustufen signifikant geringer einschätzten. Dieser Rückgang in der Selbsteinschätzung von Initiierungszeitpunkt I zu Initiierungszeitpunkt II lässt sich für den ersten Messzeitpunkt für alle Items beobachten und unterscheidet sich auch im Ausmaß kaum. Für den zweiten Messzeitpunkt liegen tendenziell die gleichen Entwicklungen vor, jedoch sind die Abweichungen zwischen den Initiierungszeitpunkten weniger groß und auch nicht für alle Items signifikant.

Überzeugungen

Am Ende eines jeden Aufgabenkomplexes wurden die Studierenden auf einer sechsstufigen Likert-Skala von $1 = \text{gar nicht überzeugt}$ bis $6 = \text{sehr überzeugt}$ im Rahmen einer weiteren Selbsteinschätzung gefragt, wie überzeugt sie davon sind, die jeweilige Aufgabe korrekt gelöst

³⁵ Für die Selbsteinschätzungen, die sich auf die Bearbeitung von Buchungs- und Transferaufgaben bezogen, erfolgte nach der jeweiligen Bearbeitung der Aufgaben noch eine dritte Selbsteinschätzung (SE171 und SE191 bzw. SE271 und SE291), die in der Folge mit *Aufstellung von Buchungssätzen II* und *Transfer auf neue Aufgaben II* gekennzeichnet werden.

³⁶ Parametrische Tests sind nach Döring & Bortz (2016) insbesondere bei großen Stichproben robust gegen die Verletzung von Testvoraussetzungen. Aus diesem Grund werden in der Folge, sofern nicht anders angegeben, parametrische Tests durchgeführt.

zu haben. In Tabelle 15 sind die Mittelwerte und Standardabweichungen der Überzeugungen zu den einzelnen Aufgabenkomplexen für beide Messzeitpunkte dargestellt.

Konsistent zu den bisher betrachteten Initiierungszeitpunkten der Selbsteinschätzung sind die Überzeugungen für den zweiten Messzeitpunkt höher als für den ersten Messzeitpunkt. Für den ersten Messzeitpunkt zeigen sich die höchsten Überzeugungswerte für die korrekte Bearbeitung der dritten Textaufgabe, die die Definition und Erläuterung des Belegprinzips zum Gegenstand hatte ($M = 3.72$, $SD = 1.47$). Alle weiteren Werte des ersten Messzeitpunktes liegen unterhalb der Skalenmitte von 3.5. Zum zweiten Messzeitpunkt sind die Studierenden besonders davon überzeugt, Buchungsaufgaben lösen zu können ($M = 4.94$, $SD = 1.15$). Auch die Überzeugung, die MC-Aufgaben ($M = 3.96$, $SD = 1.23$) und die dritte Textaufgabe ($M = 3.99$, $SD = 1.26$) korrekt gelöst zu haben, liegt jeweils über der Skalenmitte.

Performanz

Im Rahmen der Testung wurden verschiedene Aufgabenformate berücksichtigt, die unterschiedliche Wissensarten für eine adäquate Bearbeitung voraussetzen (s. Kapitel 5.2.2.1). In Tabelle 16 sind die Mittelwerte und Standardabweichungen für alle eingesetzten Aufgaben zu beiden Messzeitpunkten dargestellt.

Mit den Text- und MC-Aufgaben wurden sowohl Faktenwissen als auch konzeptuelles Wissen geprüft. Die Studierenden hatten zum ersten Messzeitpunkt noch Schwierigkeiten bei der Bearbeitung der Textaufgaben. Bei allen Textaufgaben erreichten die Studierenden nicht annähernd die Hälfte der jeweils maximal zu erreichenden Punkte. Sie konnten lediglich bei der Aufgabe zur Definition und Erläuterung des Belegprinzips (TA133) nennenswert punkten. So erhielten bei dieser Aufgabe etwa 45 % der Studierenden zwei oder mehr von vier möglichen Punkten. Gleichwohl die durchschnittlich von den Studierenden erreichten Punkte in allen Textaufgaben auf einem niedrigen Niveau sind ($M = 1.79$ Punkte, $SD = 1.47$), haben einige wenige Studierende einzelne Textaufgaben vollumfänglich gelöst (z. B. 3.1 % der Studierenden bei Aufgabe TA133). Zwar schnitten die Studierenden zum zweiten Messzeitpunkt am Ende des Semesters besser ab, jedoch erreichten sie auch dann im Durchschnitt nicht die Hälfte der maximalen Punktzahl ($M = 3.29$ Punkte, $SD = 2.24$).

Mit den MC-Aufgaben wurden die Studierenden angehalten, Buchungssätze potentiell zugrundeliegenden Sachverhalten zuzuordnen, was konzeptuelles Wissen über erfolgswirksame und –unwirksame Geschäftsvorfälle erfordert. Dadurch, dass einem Buchungssatz auch mehrere Sachverhalte zugrunde liegen können, wurde die Wahrscheinlichkeit, die richtige Lösung per

Zufall zu treffen (Ratewahrscheinlichkeit), reduziert. Im Vergleich zu den Textaufgaben sind die Lösungsquoten bei den MC-Aufgaben höher. So konnten zum ersten Messzeitpunkt drei MC-Aufgaben (MC142, MC144, MC147) von mehr als der Hälfte der Studierenden korrekt gelöst werden. Im Mittel erreichten die Studierenden bei den MC-Aufgaben von sieben möglichen Punkten 3.61 Punkte ($SD = 2.08$), wobei 30 Studierende die maximale Punktzahl erhielten. Zum zweiten Messzeitpunkt am Semesterende stieg die von den Studierenden durchschnittlich erzielte Punktzahl auf 5.10 Punkte ($SD = 1.59$). Mehr als die Hälfte der Studierenden erreichte mindestens sechs von sieben möglichen Punkten.

Für die sechs zu bearbeitenden Buchungssätze konnten die Studierenden jeweils zwei Punkte erreichen. Von den daraus resultierenden maximal 12 zu vergebenden Punkten erhielten sie zum ersten Messzeitpunkt im Durchschnitt 4.29 Punkte ($SD = 3.80$). Sie verfügten demnach zu diesem Zeitpunkt nur bedingt über das notwendige prozedurale Wissen, um Buchungssätze korrekt aufzustellen. Die Performanz bei der Bearbeitung der Buchungsaufgaben verbesserte sich zum zweiten Messzeitpunkt deutlich. Jeder vierte Studierende erreichte nun die Höchstpunktzahl. Im Mittel erhielten die Studierenden 9.75 Punkte ($SD = 2.38$).

Die Transferaufgabe konnte von den Studierenden zum ersten Messzeitpunkt kaum gelöst werden. Lediglich 12 % der Studierenden präsentierten adäquate Lösungsansätze und erhielten entsprechend zumindest Teilpunkte. Im Durchschnitt erreichten die Studierenden bei dieser Aufgabe .28 Punkte ($SD = .90$) von vier möglichen Punkten. Auch zu Semesterende blieb die Lösungsquote bei dieser Aufgabe gering. Im Mittel wurden nun .58 Punkte ($SD = 1.35$) erzielt.

In der Summe aller Aufgaben erlangten die Studierenden zum ersten Messzeitpunkt im Mittel 9.97 Punkte ($SD = 6.36$). Die Spannweite der erzielten Punkte reichte dabei von 0 bis 28 von maximal 34 zu erreichenden Punkten. Am Semesterende haben die Studierenden im Schnitt fast doppelt so viele Punkte erzielt ($M = 18.74$ Punkte, $SD = 5.13$). Die Spannweite lag zwischen 1 und 32 Punkten.

	<i>max. Punkte</i>	Messzeitpunkt I		Messzeitpunkt II	
		(N = 389)		(N = 163)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Aufgabe³⁷					
TA131: Prinzip der getrennten Kontenführung	2	.11	.38	.40	.70
TA132: Tatbestandselemente steuerbarer Umsätze	5	.22	.78	.94	1.69
TA133: Belegprinzip	4	1.47	.95	1.96	.81
Summe Textaufgaben	11	1.79	1.47	3.29	2.24
MC141: Zinsertrag	1	.47	.50	.80	.40
MC142: Privatentnahme	1	.61	.49	.83	.38
MC143: Rohstoffaufwand	1	.47	.50	.57	.50
MC144: GuV-Konto	1	.57	.50	.66	.48
MC145: Verbindlichkeiten	1	.37	.48	.63	.49
MC146: Privateinlage	1	.50	.50	.79	.41
MC147: Mietaufwand	1	.62	.49	.83	.38
Summe MC-Aufgaben	7	3.61	2.08	5.10	1.59
BS161: Mietertrag	2	1.07	.98	1.90	.42
BS162: Tilgung eines Bankkredits	2	.93	.97	1.80	.59
BS163: Abschluss „Privatentnahmekonto“	2	.50	.86	1.21	.98
BS164: Kauf von Maschinen auf Ziel	2	.47	.81	1.61	.74
BS165: Privateinlage	2	1.10	.99	1.80	.59
BS166: Verkauf von Waren	2	.22	.53	1.43	.73
Summe Buchungsaufgaben	12	4.29	3.80	9.75	2.38
FS181: Ermittlung Gewinn und Eigenkapital	4	.28	.90	.58	1.35
Summe Transferaufgabe	4	.28	.90	.58	1.35
Summe Gesamt	34	9.97	6.36	18.74	5.13

Tabelle 16: Übersicht über Performanzwerte an beiden Messzeitpunkten

³⁷ Die Itembezeichnungen beziehen sich in dieser Darstellung auf den ersten Messzeitpunkt. Für den zweiten Messzeitpunkt bilden sie sich analog (z. B. TA231, TA232, etc.).

6.1.2 Selbsteinschätzungen und Performanz der Teilstichprobe

Im Folgenden werden die Angaben und Ergebnisse der 56 Studierenden (Teilstichprobe) dargestellt, die an beiden Messzeitpunkten teilnahmen. Über die Codeabfrage konnten die Fragebögen eindeutig zugeordnet werden. Da für die Prüfung der Hypothesen 2 bis 4 auf die Angaben der Studierenden zurückgegriffen wird, die an beiden Messzeitpunkten an der Befragung teilgenommen haben, werden zunächst auch für die Teilstichprobe die deskriptiven Befunde angeführt.

Selbsteinschätzungen

Ein t-Test für abhängige Stichproben wurde durchgeführt, um zu überprüfen, ob sich die Selbsteinschätzungen vor der Bearbeitung der Aufgaben (Initiierungszeitpunkt I) zum zweiten Messzeitpunkt signifikant von jenen zum ersten Messzeitpunkt unterscheiden (s. Anhang A4). Es zeigt sich, dass die Einschätzungen, noch nicht über das nötige Verständnis zur Buchführung zu verfügen, signifikant voneinander abweichen ($t(55) = -4.23$, $p < .001$). Die Studierenden neigen zum ersten Messzeitpunkt stärker dazu, dieser Aussage zuzustimmen ($M = 2.80$, $SD = 1.29$) als zum zweiten Messzeitpunkt ($M = 1.96$, $SD = 1.03$). Signifikante Unterschiede lassen sich auch für die weiteren Items zur Selbsteinschätzung vor der Aufgabenbearbeitung beobachten. So geben Studierende zum zweiten Messzeitpunkt eher an, grundlegende Begriffe und Prinzipien wiedergeben ($M = 4.23$, $SD = 1.01$), Zusammenhänge erläutern ($M = 4.16$, $SD = 1.09$), Buchungssätze aufstellen ($M = 4.82$, $SD = .86$) sowie Transferaufgaben bearbeiten zu können ($M = 3.70$, $SD = 1.08$). Demnach sind die Selbsteinschätzungen vor der Aufgabenbearbeitung zum zweiten Messzeitpunkt durchgängig gestiegen. Dieser Befund spiegelt sich auch beim Vergleich der absoluten Werte wider (s. Abbildung 13). Dargestellt ist die jeweilige Differenz zwischen den Selbsteinschätzungen am ersten und zweiten Messzeitpunkt, wobei die Ausprägung der Selbsteinschätzung zum ersten Messzeitpunkt von jener zum zweiten Messzeitpunkt subtrahiert wurde. Positive Werte bedeuten somit, dass die Selbsteinschätzung gestiegen ist, negative Werte weisen auf einen Rückgang hin.

Nennenswert negative Ausprägungen liegen für SE22³⁸ vor. Studierende sind folglich zum zweiten Messzeitpunkt weniger der Meinung, noch nicht über das nötige Verständnis für die Buchführung zu verfügen. Aufgrund der negativen Polung dieses Items ist die Veränderung allerdings konsistent zu jenen der anderen Items zur Selbsteinschätzung. Für die weiteren Items

³⁸ SE22 = SE222 - SE122 (fehlendes Verständnis), SE23 = SE223 - SE123 (Wiedergabe grundlegender Begriffe & Prinzipien), SE24 = SE224 - SE124 (Erläuterung von Zusammenhängen), SE25 = SE225 - SE125 (Aufstellung von Buchungssätzen), SE26 = SE226 - SE126 (Transfer auf neue Aufgaben)

(SE23 bis SE26) sind die Selbsteinschätzungen mehrheitlich gestiegen. Insbesondere die Selbsteinschätzung zur korrekten Lösung von Transferaufgaben (SE26) weist zum zweiten Messzeitpunkt höhere Werte auf. Neben der gestiegenen Selbsteinschätzung bei vielen Studierenden, ist ferner auffallend, dass die Selbsteinschätzung bei einem Teil der Studierenden zu beiden Messzeitpunkten stabil ist. So liegt der Anteil der Studierenden, die ihre Selbsteinschätzung nicht änderten, zwischen 25 % für das Lösen von Transferaufgaben (SE26) und 37 % für das Aufstellen von Buchungssätzen (SE25).

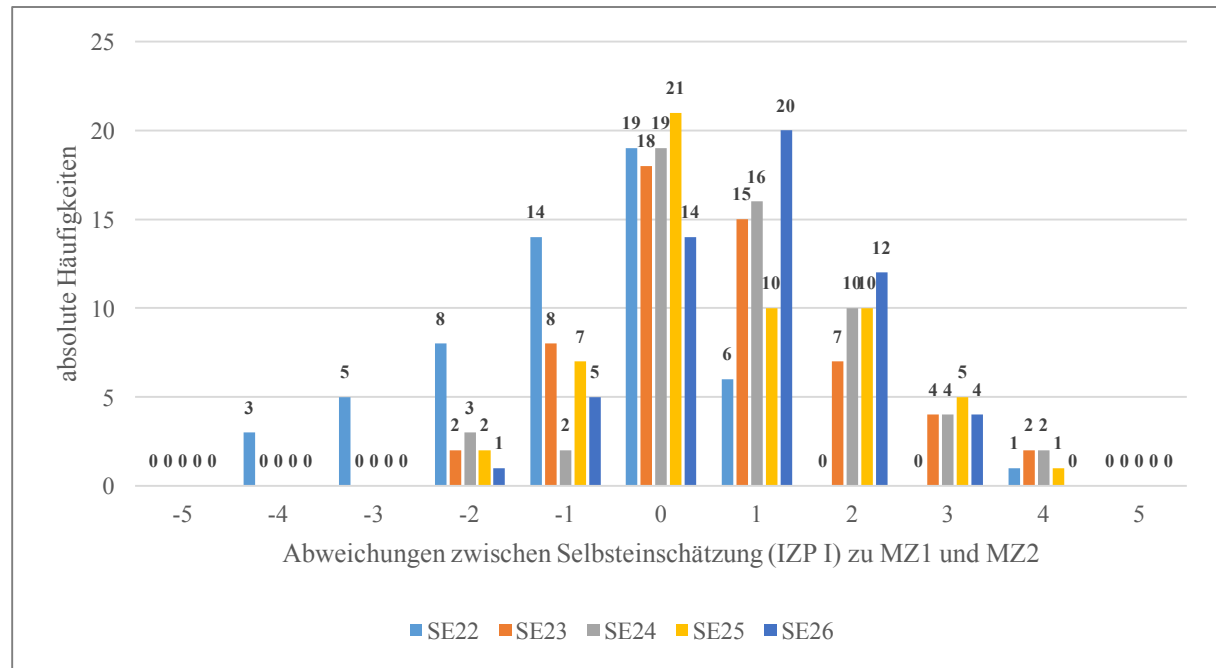


Abbildung 13: Vergleich der Selbsteinschätzungen zwischen beiden Messzeitpunkten vor der Aufgabenbearbeitung

Die höheren Selbsteinschätzungen im Verlauf des Semesters lassen sich auch für den zweiten Initiierungszeitpunkt der Selbsteinschätzung beobachten (s. Anhang A5). Die Studierenden stimmen der Aussage, noch nicht über das nötige Verständnis zur Buchführung zu verfügen, zum zweiten Messzeitpunkt signifikant weniger zu ($t(54) = -4.49$, $p < .001$) als zum ersten Messzeitpunkt. Signifikant höher wurden zum zweiten Messzeitpunkt außerdem die Aussagen zur Wiedergabe grundlegender Begriffe und Prinzipien ($t(55) = 3.80$, $p < .001$), zur Erläuterung der Zusammenhänge grundlegender Begriffe und Prinzipien ($t(55) = 6.85$, $p < .001$), zum Aufstellen von Buchungssätzen I ($t(55) = 4.97$, $p < .001$) sowie zur Bearbeitung von Transferaufgaben I ($t(55) = 6.69$, $p < .001$) bewertet.

Nach der Bearbeitung der Buchungs- und Transferaufgaben stiegen die Unterschiede in der Selbsteinschätzung zwischen den Messzeitpunkten noch an³⁹.

Hinsichtlich der absoluten Abweichungen der Selbsteinschätzungen nach der Aufgabenbearbeitung⁴⁰ ergibt sich ein ähnliches Bild wie zum ersten Initiierungszeitpunkt (s. Abbildung 14). Auffällig ist jedoch der in Teilen etwas geringere Anteil konstant gebliebener Selbsteinschätzungen und die dafür stärker ausgeprägten leicht höheren Selbsteinschätzungen.

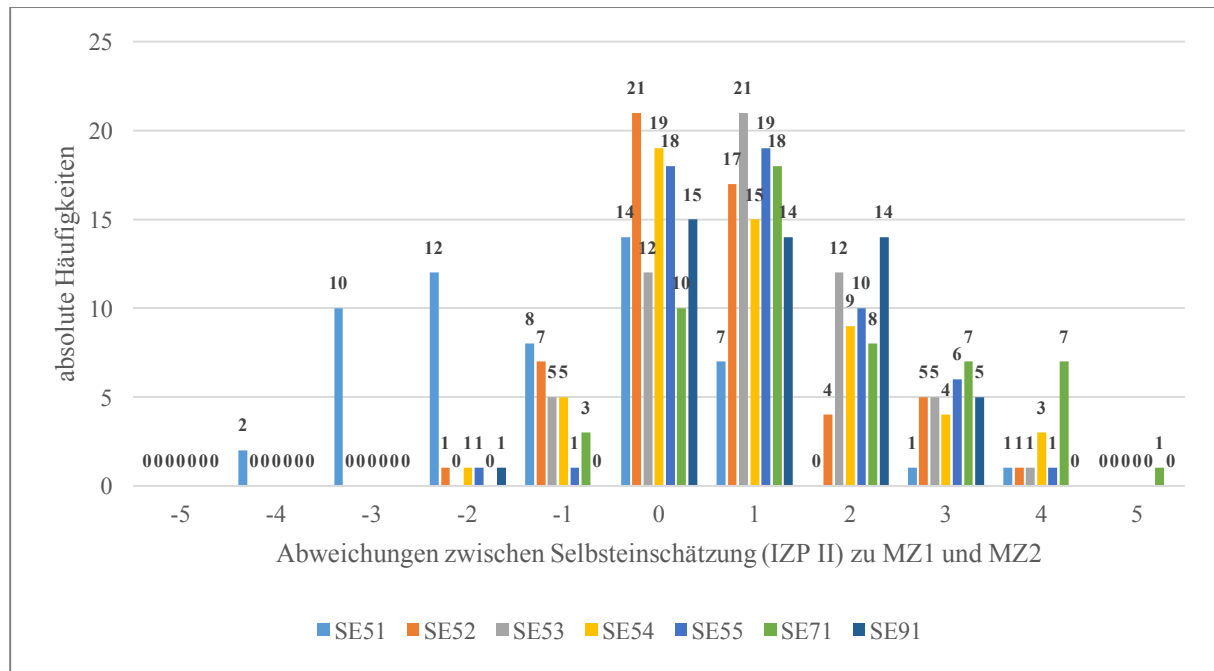


Abbildung 14: Vergleich der Selbsteinschätzungen zwischen beiden Messzeitpunkten nach der Aufgabenbearbeitung

Überzeugungen

Auch die Überzeugungen, die jeweiligen Aufgaben korrekt gelöst zu haben, sind zum zweiten Messzeitpunkt, verglichen mit dem ersten Messzeitpunkt, signifikant gestiegen (s. Anhang A6). Die einzige Ausnahme bildet die dritte Textaufgabe, bei der sich die Überzeugung nicht signifikant geändert hat ($t(48) = 1.41$, $p = .164$). Der Mittelwert ($M = 3.82$) war jedoch bereits beim ersten Messzeitpunkt auf einem vergleichsweise hohen Niveau. Am zweiten Messzeitpunkt sind die Studierenden hinsichtlich der ersten Textaufgabe ebenso signifikant überzeugter von der Richtigkeit der Antwort ($t(45) = 3.30$, $p = .002$) wie auch bei der zweiten Textaufgabe ($t(30)$

³⁹ Die Selbsteinschätzung zur Buchungs- und Transferaufgabe wurde zum einen nach der Bearbeitung der Text- und MC-Aufgaben (Buchungs-/Transferaufgabe I genannt) und zum anderen nach der Bearbeitung der Buchungs- und Transferaufgabe erfasst (Buchungs-/Transferaufgabe II genannt).

⁴⁰ SE51 = SE251 - SE151 (fehlendes Verständnis), SE52 = SE252 - SE152 (Wiedergabe grundlegender Begriffe & Prinzipien), SE53 = SE253 - SE153 (Erläuterung von Zusammenhängen), SE54 = SE254 - SE154 (Aufstellung von Buchungssätzen I), SE55 = SE255 - SE155 (Transfer auf neue Aufgaben I), SE71 = SE271 - SE171 (Aufstellung von Buchungssätzen II), SE91 = SE291 - SE191 (Transfer auf neue Aufgaben II)

= 3.56, $p = .001$). Diese Tendenz setzt sich bei den MC-Aufgaben ($t(51) = 5.68$, $p < .001$), den Buchungsaufgaben ($t(43) = 9.09$, $p < .001$) und der Transferaufgabe ($t(26) = 2.27$, $p = .032$) fort.

Die Betrachtung der absoluten Abweichungen zeigt, dass insbesondere die Überzeugungen zu den MC-Aufgaben (SE48⁴¹) und den Buchungssätzen (SE67) häufig gestiegen sind (s. Abbildung 15). Je nach Aufgabe schwankt der Anteil stabiler Überzeugungen zwischen knapp 10 % für das Lösen der Buchungsaufgaben (SE67) und 32 % für das Lösen der dritten Textaufgabe (SE33).

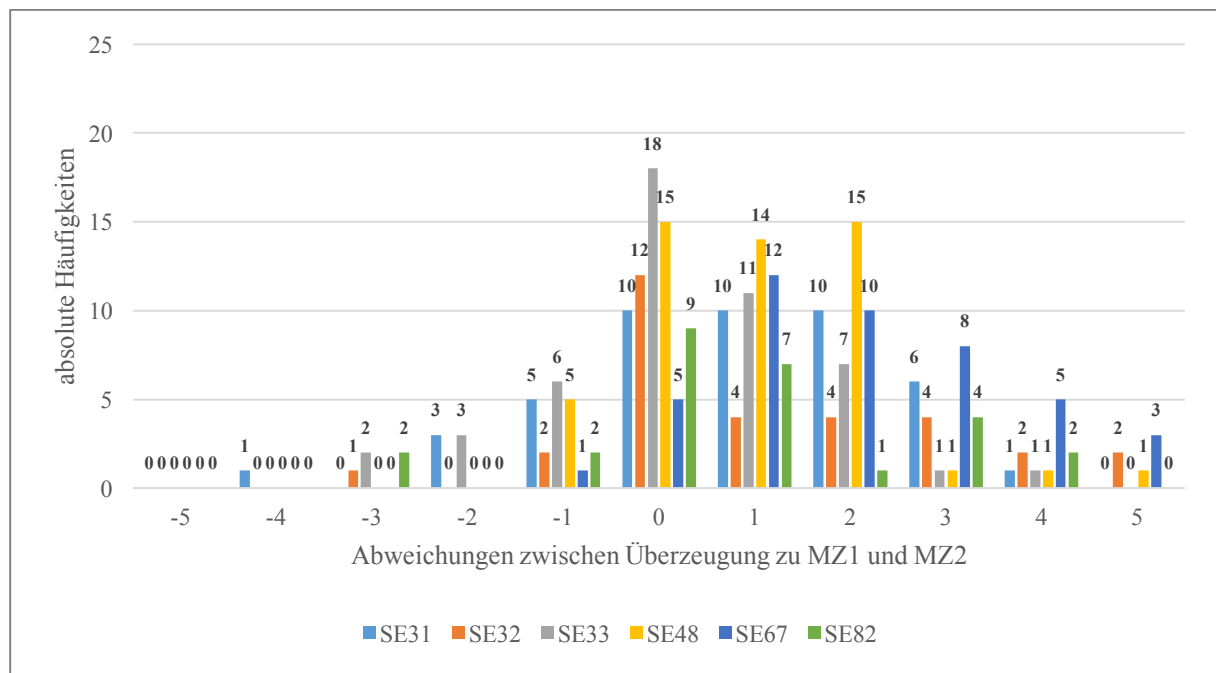


Abbildung 15: Vergleich der Überzeugungen zur korrekten Bearbeitung der Aufgaben zwischen beiden Messzeitpunkten

Performanz

Die am zweiten Messzeitpunkt erreichten Punkte in den verschiedenen Aufgaben unterscheiden sich signifikant von den am ersten Messzeitpunkt erreichten (s. Abbildung 16). Bei den Textaufgaben ($t(56) = 3.89$, $p < .001$), den MC-Aufgaben ($t(56) = 4.69$, $p < .001$), den Buchungsaufgaben ($t(56) = 9.11$, $p < .001$) und der Gesamtpunktzahl ($t(56) = 9.72$, $p < .001$) erreichten die Studierenden zum zweiten Messzeitpunkt mehr Punkte als zuvor. Dabei war der Anstieg der Performanz bei der Aufstellung der Buchungssätze besonders deutlich. Lediglich bei der Transferaufgabe lässt sich kein signifikanter Punktgewinn registrieren ($t(56) = 1.82$, $p = .075$).

⁴¹ SE31 = SE231 - SE131 (Textaufgabe I), SE32 = SE232 - SE132 (Textaufgabe II), SE33 = SE233 - SE133 (Textaufgabe III), SE48 = SE248 - SE148 (MC-Aufgaben), SE67 = SE267 - SE167 (Buchungsaufgaben), SE82 = SE282 - SE182 (Transferaufgabe)

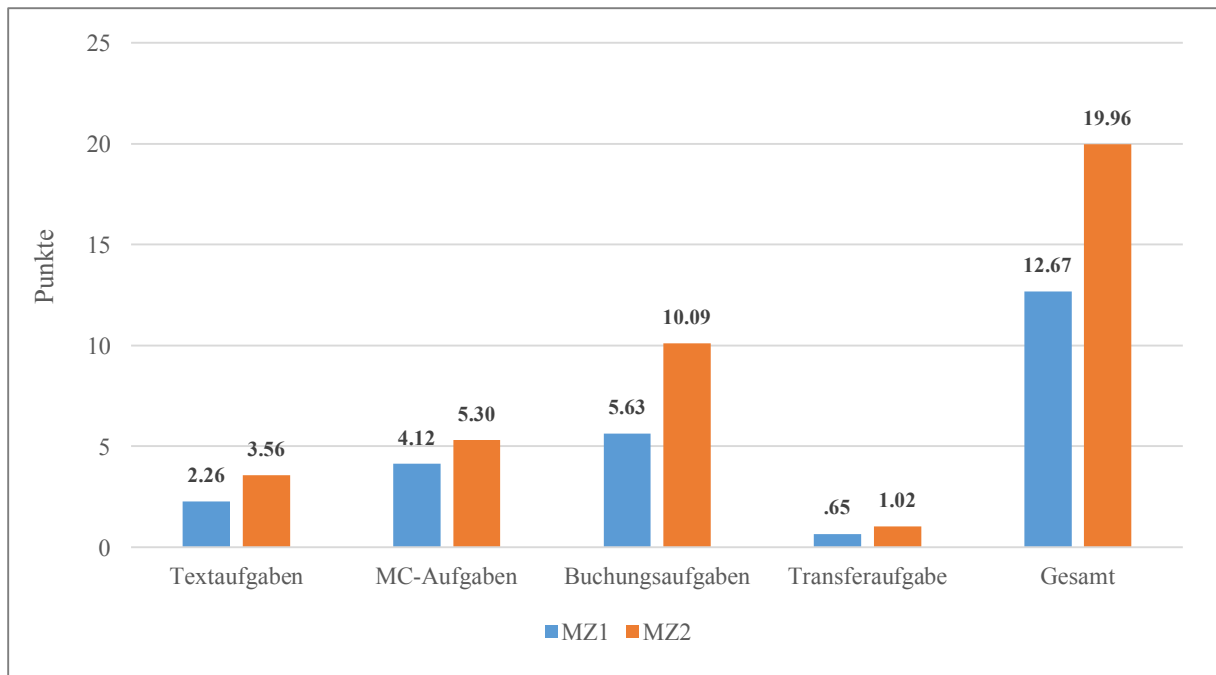


Abbildung 16: Vergleich der an beiden Messzeitpunkten im Mittel erreichten Punkte

6.2 Befunde zur Akkuratheit der Selbsteinschätzung

Nachdem die von den Studierenden angegebenen Selbsteinschätzungen und die Performanz zunächst auf überwiegend deskriptiver Ebene dargestellt wurden, werden in diesem und den folgenden Unterkapiteln die Ergebnisse hypothesenbezogen beschrieben. Die ersten beiden Hypothesen beziehen sich auf die Akkuratheit der Selbsteinschätzung. Zur Beantwortung der Hypothesen wird im Folgenden die Akkuratheit der Selbsteinschätzung für alle Initiierungszeitpunkte der Selbsteinschätzung an beiden Messzeitpunkten analysiert.

- H_{1a} Studierende im wirtschaftswissenschaftlichen Kontext weisen überwiegend akkurate Selbsteinschätzungen auf, d. h. die Studierenden weichen in ihren Selbsteinschätzungen nicht mehr als einen Skalenpunkt von der tatsächlichen Performanz ab.
- H_{1b} Es wird erwartet, dass die Akkuratheit der Selbsteinschätzung zum zweiten Initiierungs- und Messzeitpunkt zunimmt (Maki et al., 1994, Kostons et al., 2012).

6.2.1 Ausprägung akkurater Selbsteinschätzungen an Messzeitpunkt I

Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)

Die Selbsteinschätzung vor der Aufgabenbearbeitung erfolgte einerseits über die direkte Angabe eines Niveaus, das dem aktuellen Wissens- und Könnensstand der Studierenden am ehesten entspricht (SE121) und andererseits durch die Einschätzung des eigenen Wissens und Könnens zu ebendiesen Niveaustufen auf einer sechsstufigen Likert-Skala (SE122 bis SE126).

In einem ersten Analyseschritt wird die Performanz bezüglich der Bearbeitung der Aufgabenstellungen in Abhängigkeit von der zu Beginn der Befragung angegebenen Niveaustufe des aktuellen Wissens- und Könnensstandes (SE121) dargestellt (s. Anhang A7). Alle Studierende, die sich auf derselben Niveaustufe einschätzten, werden zu jeweils einer Gruppe zusammengefasst. Anschließend wird das arithmetische Mittel der in den einzelnen Aufgabenbereichen erreichten Punkte für jede Gruppe berechnet. Da die Voraussetzungen zur Durchführung einer Varianzanalyse nicht vollständig erfüllt sind (fehlende Normalverteilung), wurde zur Signifikanzprüfung der Performanzunterschiede bezüglich der verschiedenen Aufgaben in Abhängigkeit von der Selbsteinschätzung (SE121) auf den Kruskal-Wallis-Test als nicht-parametrisches Verfahren zurückgegriffen. Die Ergebnisse des Kruskal-Wallis-Tests zeigen, dass für die Textaufgaben ($\chi^2 = 40.84$, $p < .01$), die MC-Aufgaben ($\chi^2 = 68.00$, $p < .01$), die Buchungsaufgaben ($\chi^2 = 83.70$, $p < .01$) und die Transferaufgabe ($\chi^2 = 11.23$, $p = .024$) signifikante Unterschiede vorliegen. Auch die Betrachtung der Summe der in allen Aufgaben erzielten Punkte bestätigt

den Befund ($\chi^2 = 97.81$, $p < .01$). Es zeigt sich, dass die Studierenden, die glauben, noch nicht über das Verständnis für die Grundlagen der Buchführung zu verfügen ($n = 88$ Studierende), bei allen Aufgaben schlechter abschneiden ($M = 4.43$ Punkte, $SD = 3.99$) als die übrigen Studierenden. Dies deutet darauf hin, dass Studierende mit noch geringen Kenntnissen bezüglich der Inhalte des externen Rechnungswesens sich dieser auch bewusst zu sein scheinen. Dagegen erreichen die Studierenden, die sich bereits in der Lage sehen, ihr Wissen auf neue Aufgaben anzuwenden ($n = 24$ Studierende), die höchsten Punktzahlen in der Summe aller Aufgaben ($M = 14.21$ Punkte, $SD = 5.90$). Gleichwohl waren auch sie kaum in der Lage, die Transferaufgabe (FS181) zu lösen ($M = .58$ Punkte, $SD = 1.38$). Dies ist ein Indiz dafür, dass die Selbsteinschätzung bezogen auf die jeweilige Wissensart und Aufgabenstellung nur bedingt akkurat zu sein scheint. Diese Tendenz zeigt sich auch bei der Betrachtung der weiteren Niveaustufen. Studierende, die angaben, grundlegende Konzepte und Prinzipien wiedergeben zu können, erreichten im Mittel nur 2.23 Punkte ($SD = 1.56$) bei der Bearbeitung der Textaufgaben. Auch für die MC-Aufgaben ($M = 4.35$, $SD = 1.63$) und die Buchungsaufgaben ($M = 5.27$, $SD = 3.64$) lassen die Performanzwerte noch nicht darauf schließen, dass die Studierenden sich einer Niveaustufe zutreffend zugeordnet haben. Entsprechend sind die Korrelationen zwischen der anfänglichen Selbsteinschätzung (SE121) und der gezeigten Performanz meist lediglich schwach ausgeprägt. Für das Lösen von Transferaufgaben ergab sich lediglich ein Zusammenhang von $r = .140$ ($p = .008$). Dagegen sind die Korrelationen bei der Bearbeitung von MC-Aufgaben ($r = .364$, $p < .001$) und Buchungsaufgaben ($r = .441$, $p < .001$) etwas stärker.

Zur weiteren Bestimmung der Akkuratheit der Selbsteinschätzung wurde die Zahl der maximal zu erreichenden Punkte in Text-, MC-, Buchungsaufgaben und Transferaufgaben zunächst jeweils auf einen Wert von fünf normiert. Außerdem wurden die auf der sechsstufigen Likert-Skala vorgenommenen Selbsteinschätzungen auf eine Skala von $0 = \text{gar nicht zutreffend}$ bis $5 = \text{sehr zutreffend}$ recodiert, um anschließend die berechneten normierten Punktzahlen den recodierten Selbsteinschätzungen auf der sechsstufigen Likert-Skala gegenüberstellen zu können, indem die erreichte normierte Punktzahl von der Ausprägung der jeweiligen recodierten Selbsteinschätzung subtrahiert wird. Ist der so ermittelte Wert positiv, liegt eine Überschätzung auf Seiten des Studierenden vor. Ein Wert von Null spricht für eine akkurate Selbsteinschätzung und ein negativer Wert legt nahe, dass Studierende sich unterschätzen. In der Folge wird in Anlehnung an Coutts & Rogers (1999) davon ausgegangen, dass ein Wert von -1 oder darunter eine bedeutsame Unterschätzung, ein Wert von 1 oder darüber eine bedeutsame Überschätzung sowie Werte zwischen -1 und 1 eine akkurate Selbsteinschätzung darstellen. Die Berechnungen

wurden für beide Mess- und für alle Initiierungszeitpunkte der Selbsteinschätzung durchgeführt. Darüber hinaus werden die durchschnittlichen betragsmäßigen Abweichungen zwischen recodierter Selbsteinschätzung und erreichter normierter Punktzahl über alle Studierenden hinweg sowie die Korrelationen als weitere Akkuratheitsmaße der Selbsteinschätzung angegeben.

Für die Selbsteinschätzungen vor der Aufgabenbearbeitung ergibt sich hinsichtlich der Akkuratheit ein ambivalentes Bild (s. Abbildung 17). Mit Ausnahme der Transferaufgabe (50.7 % der Studierenden mit akkurater Selbsteinschätzung) liegen bei den weiteren Aufgabenformaten mehrheitlich inakkurate Selbsteinschätzungen vor. Bei der Bearbeitung der MC-Aufgaben neigen die Studierenden eher zur Unterschätzung (37.4 % der Studierenden), wohingegen bei Text- und Buchungsaufgaben eher höhere Selbsteinschätzungen im Vergleich zur tatsächlich realisierten Performanz auftreten (68.8 % bzw. 49.5 % der Studierenden). Das Ausmaß der Fehleinschätzung ist betragsmäßig im Mittel für die Textaufgaben am höchsten ($M = 1.82$ Skalenpunkte, $SD = 1.19$). Der niedrigste Mittelwert der betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz liegt für die MC-Aufgaben vor ($M = 1.30$ Skalenpunkte, $SD = 0.94$). Die Differenzen bei den Buchungsaufgaben betragen betragsmäßig im Mittel 1.55 Skalenpunkte ($SD = 1.18$) und für die Transferaufgabe 1.67 Skalenpunkte ($SD = 1.31$).

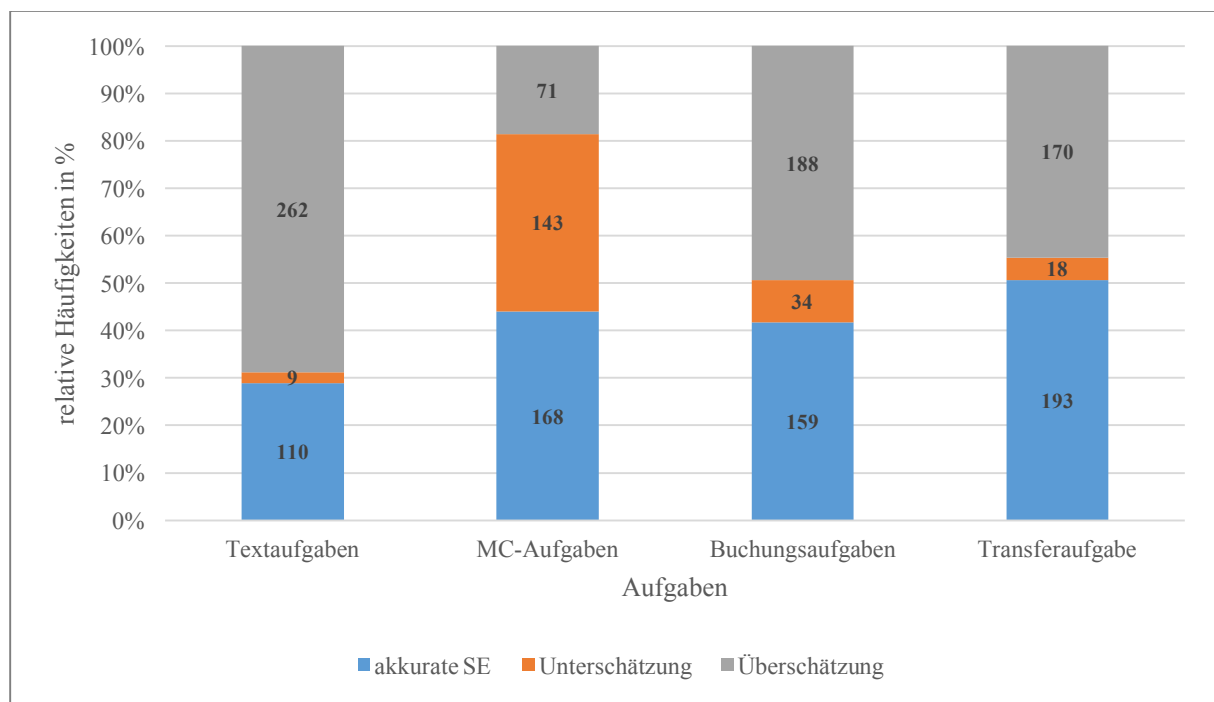


Abbildung 17: Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt I ($381 \leq n \leq 382$)

Zuletzt wurden die Korrelationen nach Pearson zwischen den Selbsteinschätzungen vor der Aufgabenbearbeitung und der jeweiligen Performanz als weiteres Maß der Akkuratheit berechnet (s. Anhang A8). Sie betragen für die Textaufgaben $r = .195$ ($p < .001$), die MC-Aufgaben $r = .371$ ($p < .001$), die Buchungsaufgaben $r = .496$ ($p < .001$) und die Transferaufgabe $r = .218$ ($p < .001$).

Die aufgezeigten Ergebnisse sprechen insgesamt dafür, dass die Selbsteinschätzungen der Studierenden vor der Aufgabenbearbeitung in einigen Teilen wenig akkurat sind. Die gleichen Berechnungen werden im Folgenden für die Selbsteinschätzungen nach der Aufgabenbearbeitung durchgeführt.

Akkuratheit der Selbsteinschätzung nach Bearbeitung der Aufgaben (Initiierungszeitpunkt II)

Die Güte der Selbsteinschätzungen nach der Aufgabenbearbeitung ist im Vergleich zu den Selbsteinschätzungen vor der Aufgabenbearbeitung in vielen Fällen gestiegen (s. Abbildung 18). Das Abschneiden sowohl bei den Buchungsaufgaben (65.2 % der Studierenden) als auch der Transferaufgabe (69.0 % der Studierenden) wurde überwiegend akkurat eingeschätzt. Dagegen hat sich die Tendenz zur Unterschätzung bei den MC-Aufgaben verstärkt (49.3 % der Studierenden). Die Akkuratheit der Selbsteinschätzung zum Abschneiden bei den Textaufgaben hat sich zwar verbessert, sie weist jedoch noch immer einen hohen Anteil an Studierenden mit Überschätzungstendenzen auf (47.2 % der Studierenden). Analog hierzu sind auch die Mittelwerte der betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz gesunken. Sie betragen zum zweiten Initiierungszeitpunkt der Selbsteinschätzung im Mittel noch zwischen .92 Skalenpunkte ($SD = .85$) für die Buchungsaufgaben II und 1.47 Skalenpunkte ($SD = 1.10$) für die MC-Aufgaben. Auch die höheren Korrelationen zwischen Selbsteinschätzung und Performanz weisen auf eine akkuratere Selbsteinschätzung nach der Aufgabenbearbeitung hin (s. Anhang A9). Sie betragen nun für die Textaufgaben $r = .269$ ($p < .001$), die MC-Aufgaben $r = .307$ ($p < .001$), die Buchungsaufgaben I $r = .564$ ($p < .001$), die Buchungsaufgaben II $r = .667$ ($p < .001$) sowie für die Transferaufgabe I $r = .254$ ($p < .001$) und Transferaufgabe II $r = .303$ ($p < .001$).

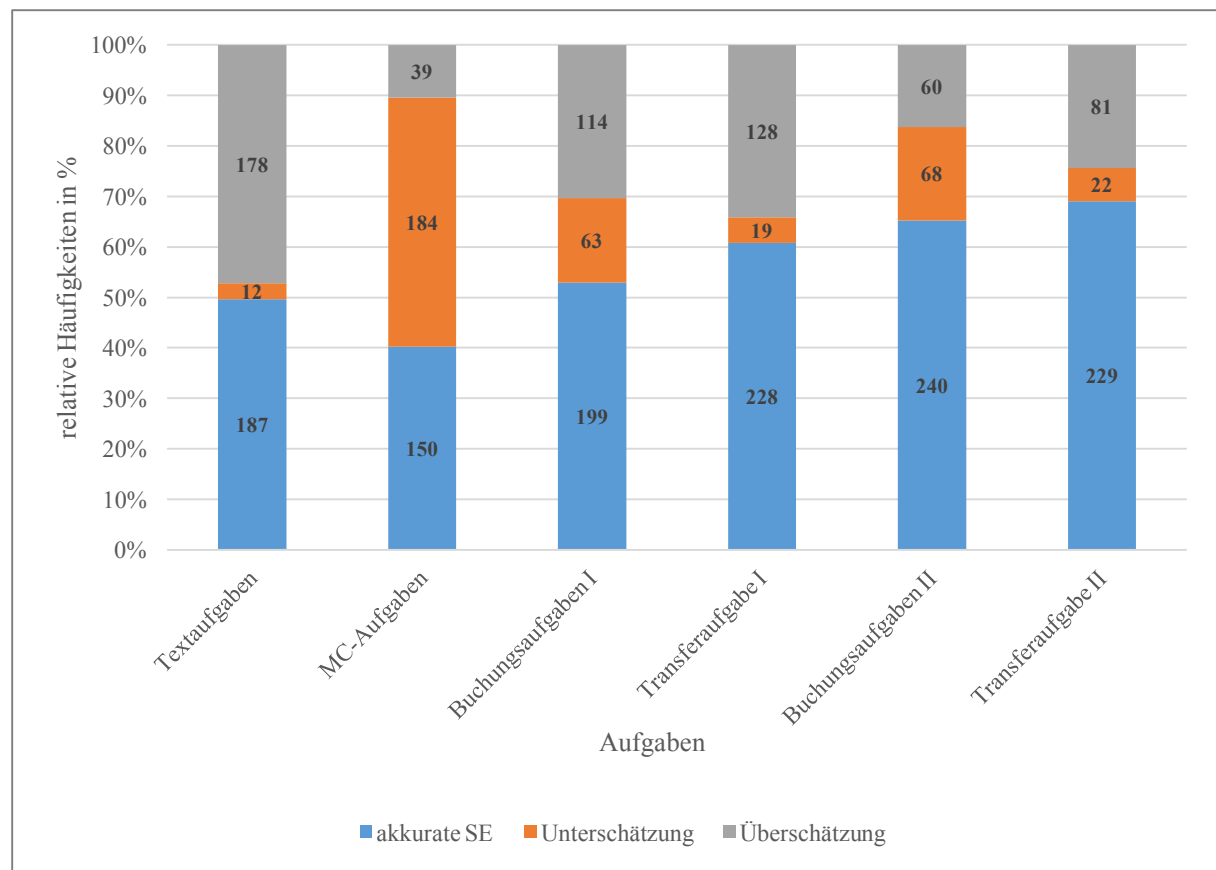


Abbildung 18: Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt I ($332 \leq n \leq 377$)

Akkuratheit der Überzeugungen

Die Überzeugung, die jeweilige Aufgabe korrekt gelöst zu haben, wurde auf Basis einer sechsstufigen Likert-Skala erfasst. Die Bestimmung der Akkuratheit der Überzeugungen erfolgt analog zur Bestimmung der Akkuratheit der Selbsteinschätzungen vor und nach der Aufgabenbearbeitung. In Abbildung 19 wird die Verteilung von akkurater Selbsteinschätzung, Unterschätzung und Überschätzung für die einzelnen Aufgaben dargestellt. Auffällig ist, dass der Anteil an akkuraten Selbsteinschätzungen für Textaufgabe I (58.8 % der Studierenden), Textaufgabe II (84.3 % der Studierenden), die Buchungsaufgaben (56.5 % der Studierenden) und die Transferaufgabe (72.5 % der Studierenden) recht hoch ausfällt. Lediglich bei der Bearbeitung der MC-Aufgaben weisen die Studierenden mehrheitlich eine Tendenz zur Unterschätzung auf (51.0 % der Studierenden). Das Ausmaß der betragsmäßigen Abweichungen zwischen Überzeugung und Performanz ist entsprechend für Textaufgabe II am geringsten ($M = .49$ Skalenpunkte, $SD = .99$) und für Textaufgabe III ($M = 1.44$ Skalenpunkte, $SD = 1.06$) und Textaufgabe I ($M = 1.35$ Skalenpunkte, $SD = 1.33$) am höchsten. Die Korrelationen zwischen Überzeugung und Performanz liegen zwischen $r = .198$ ($p < .001$) für die erste Textaufgabe und $r = .601$ ($p < .001$) für die Buchungsaufgaben (s. Anhang A10).

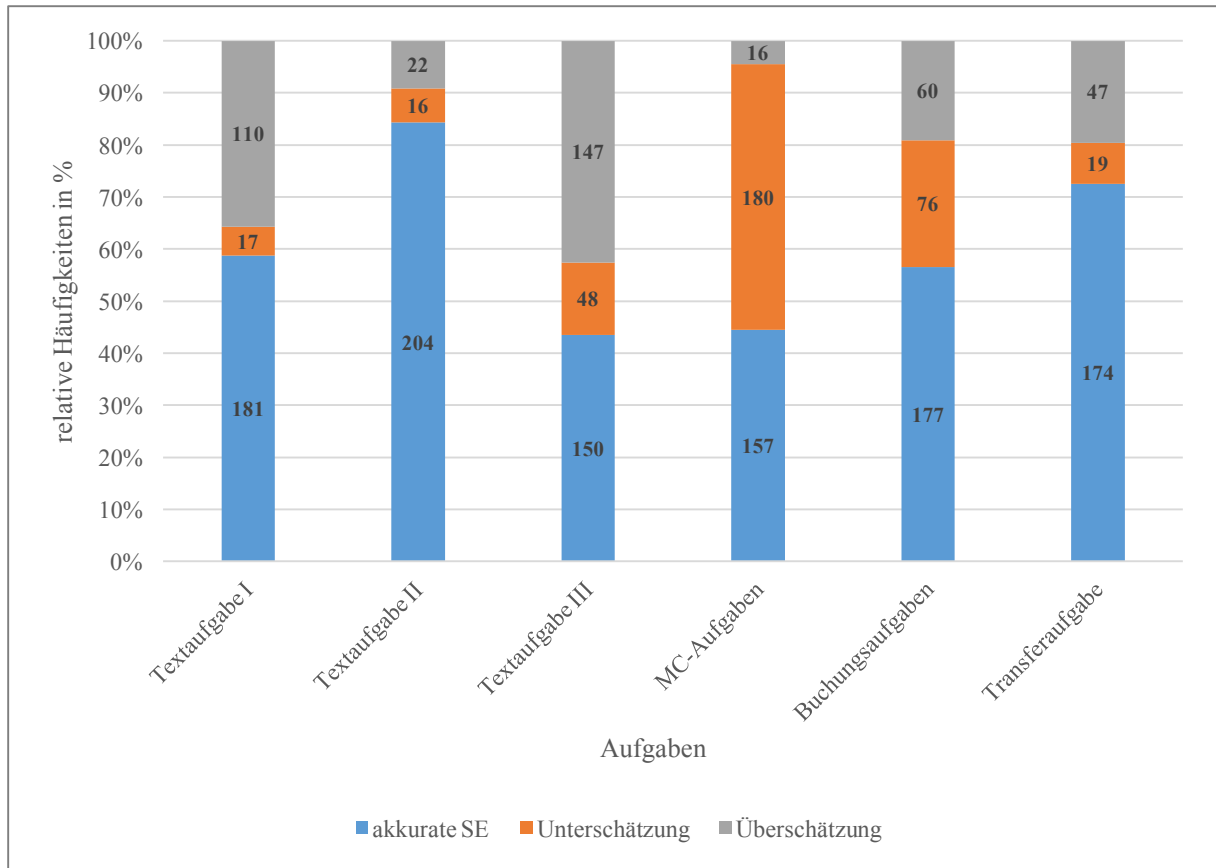


Abbildung 19: Akkuratheit der Überzeugungen an Messzeitpunkt I ($240 \leq n \leq 353$)

6.2.2 Ausprägung akkurater Selbsteinschätzungen an Messzeitpunkt II

Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)

Um die Akkuratheit der Selbsteinschätzung eruieren zu können, erfolgt im ersten Schritt wie bereits bei den Berechnungen zum ersten Messzeitpunkt der Abgleich zwischen erreichter Gesamtpunktzahl und dem zu Beginn eingeschätzten Niveau des eigenen Wissensstandes (SE221).

Erneut wurde über die Verwendung des Kruskal-Wallis-Tests zunächst überprüft, ob signifikante Unterschiede in den Performanzwerten in Abhängigkeit von der Selbsteinschätzung (SE221) vorliegen. Zeigten sich zum ersten Messzeitpunkt noch für alle Aufgaben signifikante Unterschiede, ergeben sich solche zum zweiten Messzeitpunkt nur noch für die in den MC-Aufgaben ($\chi^2 = 10.53$, $p = .032$), den Buchungsaufgaben ($\chi^2 = 19.92$, $p < .01$) und den in allen Aufgaben ($\chi^2 = 20.91$, $p < .01$) erreichten Punkte.

Die Gesamtpunktzahl (s. Anhang A11) ist für die Studierenden am höchsten, die zu Beginn angaben, bereits in der Lage zu sein, Transferaufgaben bearbeiten zu können ($n = 36$ Studierende). Sie erreichten im Durchschnitt 21.33 Punkte ($SD = 4.27$) und somit über acht Punkte mehr als jene, die die Auffassung vertraten, noch nicht über das nötige Verständnis der Buchführung zu verfügen ($n = 4$ Studierende). Auffällig ist ferner, dass die Lernenden, die glaubten, über das nötige konzeptuelle Verständnis zu verfügen ($n = 33$ Studierende), eine höhere Punktzahl erreichten ($M = 18.88$, $SD = 4.58$) als jene, die sich in der Lage sahen, Buchungssätze aufzustellen ($n = 53$ Studierende) und somit das entsprechende prozedurale Wissen besitzen müssten ($M = 18.11$, $SD = 4.81$). Selbst bei den Aufgaben, die das Aufstellen von Buchungssätzen zum Gegenstand hatten, liegt dieses Punkteverhältnis vor ($M = 9.94$ vs. $M = 9.45$). Insgesamt legen die Ergebnisse die Folgerung nahe, dass die Selbsteinschätzungen hinsichtlich des Aufstellens von Buchungssätzen verhältnismäßig akkurat sind. So liegen die erreichten Punkte bei jenen Studierenden, die meinten, über konzeptuelles und prozedurales Wissen zu verfügen bzw. Transferaufgaben lösen zu können, im Mittel zwischen 9.45 Punkten ($SD = 2.45$) und 10.89 Punkten ($SD = 1.21$) von max. 12 zu erreichenden Punkten. Damit konnten die Studierenden im Durchschnitt bei diesen Aufgaben mindestens 75 % der Gesamtpunkte erreichen. Weniger akkurat erscheinen die Einschätzungen, bereits Transferaufgaben bearbeiten zu können. Probanden, die dieser Auffassung waren, lösten die Transferaufgabe nur selten und erreichten 1.03 Punkte von max. 4 Punkten. Diese Tendenz zeigt sich auch bei der Betrachtung der Korrelationen zwischen Selbsteinschätzung (SE221) und der gezeigten Performanz, die nur

für das Lösen der Buchungsaufgaben einen schwachen signifikanten Zusammenhang aufweist ($r = .324$, $p < .001$).

Auch die weiteren Berechnungen zur Akkuratheit der Selbsteinschätzungen erfolgen äquivalent zu jenen zum ersten Messzeitpunkt. In Abbildung 20 sind die Akkuratheiten der Selbsteinschätzungen vor der Aufgabenbearbeitung für die verschiedenen Aufgabenformate dargestellt. Auffällig ist, dass für die Textaufgaben (72.2 % der Studierenden) und die Transferaufgabe (67.3 % der Studierenden) häufig Überschätzungen des eigenen Wissens- und Könnensstandes vorliegen. Dahingegen überwiegen bei den MC-Aufgaben Unterschätzungen (43.8 % der Studierenden). Lediglich für die Buchungsaufgaben liegen mehrheitlich akkurate Selbsteinschätzungen vor (51.9 % der Studierenden), gleichwohl auch hier ein beträchtlicher Teil der Studierenden zu Fehleinschätzungen neigt. Entsprechend unterschiedlich fallen die betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz im Mittel aus. Sie reichen von 1.04 Skalenpunkte ($SD = .95$) für die Buchungsaufgaben bis 2.41 Skalenpunkte ($SD = 1.20$) für die Transferaufgabe. Damit sind die Abweichungen abgesehen von den Buchungsaufgaben zum zweiten Messzeitpunkt höher als beim ersten Messzeitpunkt. Die Betrachtung der Korrelationen zwischen Selbsteinschätzung vor der Aufgabenbearbeitung und Performanz bestätigt die Tendenz inakkuraterer Selbsteinschätzungen zum zweiten Messzeitpunkt (s. Anhang A12). So liegen sie für die Textaufgaben bei $r = .228$ ($p = .004$), die MC-Aufgaben bei $r = .246$ ($p = .002$), die Buchungsaufgaben bei $r = .141$ ($p = .074$) und die Transferaufgabe bei $r = .131$ ($p = .096$) und somit niedriger als noch beim ersten Messzeitpunkt (vgl. Kapitel 6.2.1).

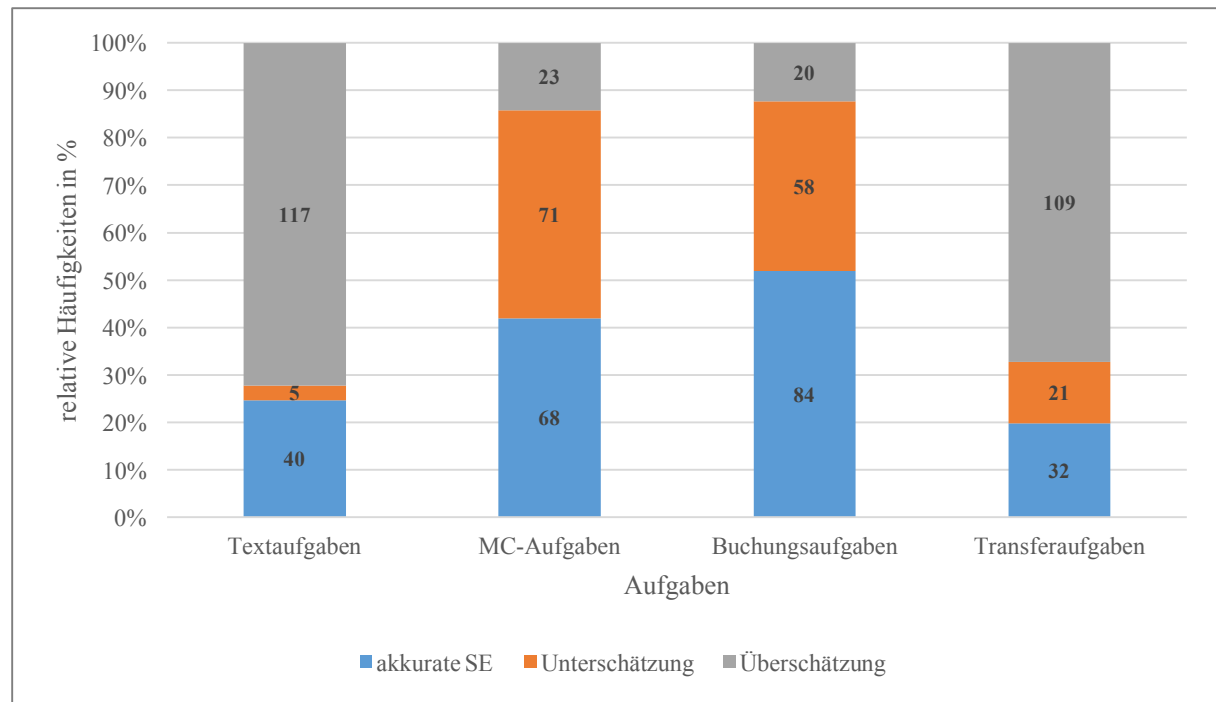


Abbildung 20: Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt II ($N = 162$)

Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung (Initiierungszeitpunkt II)

Zum zweiten Initiierungszeitpunkt der Selbsteinschätzungen ist der Anteil akkurater Selbsteinschätzungen für Textaufgaben im Vergleich zum ersten Initiierungszeitpunkt gestiegen (38.7 % der Studierenden). Dennoch neigt die Mehrheit der Studierenden auch nach der Aufgabenbearbeitung zur Überschätzung (55.2 % der Studierenden) des eigenen Wissens und Könnens. Im Gegensatz zu den Textaufgaben ist die Akkuratheit der Selbsteinschätzung für die MC-Aufgaben gesunken (31.9 % der Studierenden mit akkurater Selbsteinschätzung). Dies lässt sich auf den Anstieg zu niedrig eingeschätzten Wissens (60.1 % der Studierenden) zurückführen. Hinsichtlich der Selbsteinschätzungen zu den Buchungsaufgaben und Transferaufgaben lassen sich kaum Änderungen ausmachen (s. Abbildung 21). Die durchschnittlichen absoluten Abweichungen liegen betragsmäßig zwischen .84 Skalenpunkte ($SD = .67$) für die Selbsteinschätzung nach Bearbeitung der Buchungsaufgaben und 2.03 Skalenpunkte ($SD = 1.20$) für die Selbsteinschätzung nach der Bearbeitung der Transferaufgabe. Die Korrelationen sind für die Selbsteinschätzung zum Abschneiden bei Transferaufgaben entsprechend niedrig ($r = .105$, $p = .183$ für Transferaufgaben I und $r = .307$, $p < .001$ für Transferaufgaben II). Stärkere Zusammenhänge zwischen Selbsteinschätzung und Performanz liegen für die Buchungsaufgaben vor ($r = .568$, $p < .001$). Die weiteren Korrelationen sind in Anhang A13 dargestellt.

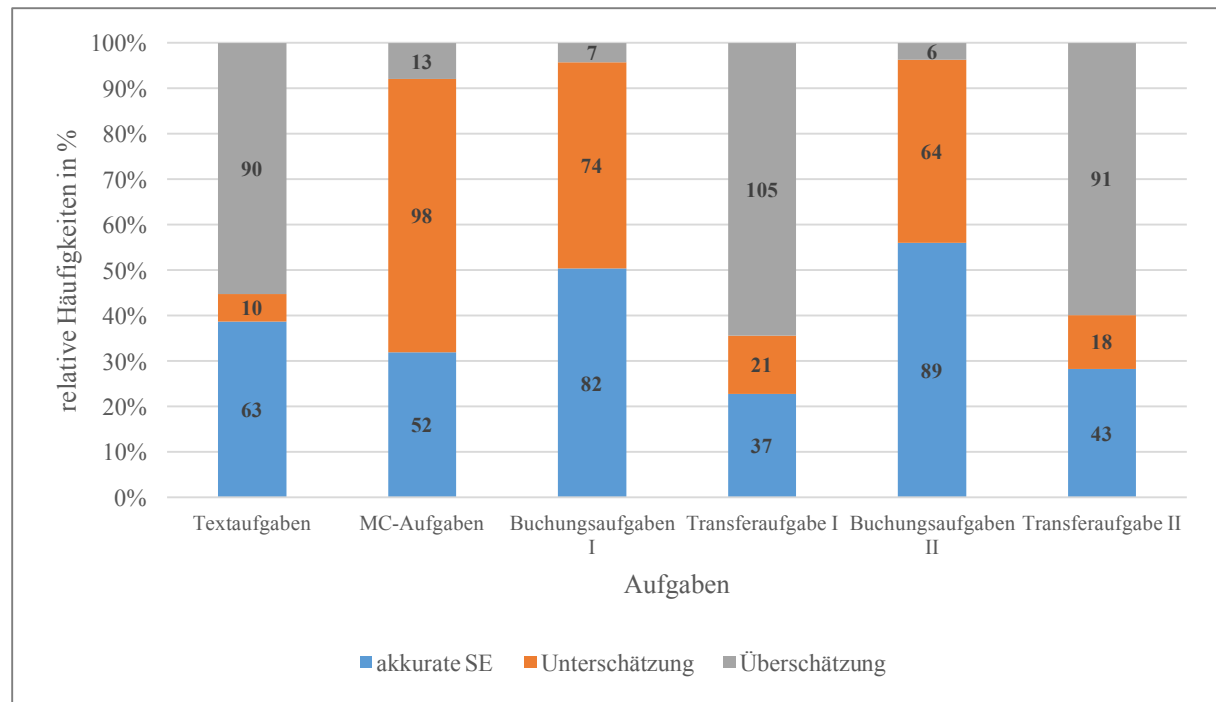


Abbildung 21: Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt II ($159 \leq n \leq 163$)

Akkuratheit der Überzeugungen

Die Überzeugung, die Aufgabe richtig gelöst zu haben, ist je nach Aufgabe als recht akkurat zu beurteilen. Die Textaufgaben II und III sowie die Buchungsaufgaben weisen mehr akkurate Überzeugungen auf als inakkurate (s. Abbildung 22). Dagegen überwiegen bei den MC-Aufgaben Unterschätzungen der eigenen Performanz (45.5 % der Studierenden). Im Mittel bewegen sich die betragsmäßigen Abweichungen zwischen Überzeugung und tatsächlich realisierter Punktzahl zwischen .74 Skalenpunkte ($SD = .70$) für die Buchungsaufgaben und 1.70 Skalenpunkte ($SD = 1.37$) für Textaufgabe I, bei der die meisten Überschätzungen vorlagen. Die Korrelationen zwischen angegebener Überzeugung und erreichter Punktzahl waren für die Textaufgabe II ($r = .696$, $p < .001$) und den Buchungsaufgaben ($r = .557$, $p < .001$) am höchsten (s. Anhang A14).

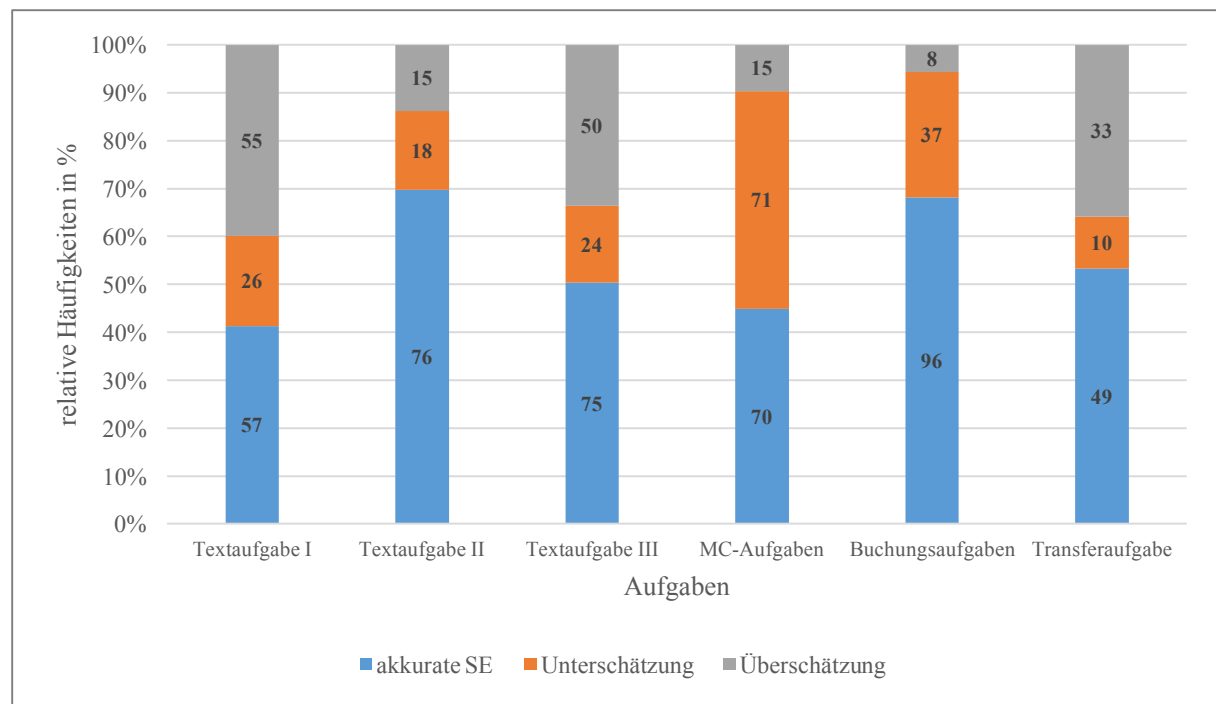


Abbildung 22: Akkuratheit der Überzeugungen an Messzeitpunkt II ($92 \leq n \leq 156$)

Nachdem die Akkuratheit der Selbsteinschätzung skizziert wurde, werden in einem Exkurs als Zwischenschritt vor der Prüfung des möglichen Zusammenhangs zwischen Akkuratheit der Selbsteinschätzung und Performanz, mögliche Einflussfaktoren auf die Akkuratheit der Selbsteinschätzung untersucht.

6.2.3 Exkurs: Einflussfaktoren auf die Akkuratheit der Selbsteinschätzungen

Geschlecht

Für den ersten Messzeitpunkt zeigen sich zum ersten Initiierungszeitpunkt der Selbsteinschätzung keine signifikanten Unterschiede zwischen weiblichen und männlichen Studierenden (s. Anhang A15). Erst für den zweiten Initiierungszeitpunkt existieren signifikante Unterschiede in der Akkuratheit der Selbsteinschätzung. Sowohl für die Buchungsaufgaben II ($M = 1.15$, $SD = .89$) als auch für die Transferaufgabe II ($M = 1.42$, $SD = 1.26$) sind die Selbsteinschätzungen der männlichen Studierenden inakkurater als die ihrer weiblichen Kommilitonen ($M = .82$, $SD = .79$ für Buchungsaufgaben II bzw. $M = .93$, $SD = 1.09$ für Transferaufgabe II). Bei den Überzeugungen, die jeweilige Aufgabe korrekt gelöst zu haben, liegen signifikante Abweichungen für die Überzeugungen zu den Buchungsaufgaben vor ($t(295) = -2.682$, $p = .008$). Demnach sind männliche Studierende ($M = 1.27$, $SD = 1.07$) in ihren Überzeugungen inakkurater als weibliche Studierende ($M = .93$, $SD = .85$).

Auch am zweiten Messzeitpunkt liegen für den ersten Initiierungszeitpunkt der Selbsteinschätzung keine signifikanten Unterschiede in der Akkuratheit der Selbsteinschätzung für männliche und weibliche Studierende vor (s. Anhang A16). Signifikante Unterschiede bestehen nur am zweiten Initiierungszeitpunkt für die Buchungsaufgaben I ($t(111) = 2.557, p = .012$). Hier sind die männlichen Studierenden in ihren Selbsteinschätzungen akkurater als die weiblichen Studierenden.

Ein ambivalentes Bild ergibt auch die Betrachtung der Akkuratheit der Überzeugungen. Am ersten Messzeitpunkt sind die Überzeugungen hinsichtlich der korrekten Bearbeitung der Buchungsaufgaben für die weiblichen Studierenden signifikant akkurater als für die männlichen Studierenden, wohingegen die männlichen Studierenden am zweiten Messzeitpunkt akkuratere Selbsteinschätzungen für die Bearbeitung der MC-Aufgaben aufweisen ($t(105) = 2.321, p = .022$).

Insgesamt liegen nur wenige signifikante Unterschiede für männliche und weibliche Studierende in der Akkuratheit der Selbsteinschätzung vor. Die wenigen Abweichungen zeigen darüber hinaus keine einheitliche Tendenz. Das Geschlecht war bei der Befragung offenbar kein Einflussfaktor auf die Akkuratheit der Selbsteinschätzung.

Hochschulzugangsberechtigung (HZB)

Es wurde untersucht, inwiefern die Akkuratheit der Selbsteinschätzung vom Typ der HZB abhängt, indem die Studierenden mit einem Abschluss am allgemeinbildenden Gymnasium jenen an anderen Schulformen erworbenen Berechtigungen (z. B. berufsbildende Gymnasien) gegenübergestellt wurden. Die Akkuratheit der Selbsteinschätzung hängt nur in wenigen Fällen vom Typ der HZB ab. Zum ersten Messzeitpunkt liegen signifikante Mittelwertunterschiede in der Akkuratheit der Selbsteinschätzung für die Einschätzung des Abschneidens bei Text- ($t(361) = -2.420, p = .016$) und Transferaufgaben ($t(361) = -3.750, p < .001$) zum ersten Initiierungszeitpunkt vor (s. Anhang A17). Studierende, die ihre HZB am allgemeinbildenden Gymnasium erworben haben, übertreffen demnach die Studierenden mit sonstiger HZB in der Akkuratheit der Selbsteinschätzung. Diese Tendenz setzt sich zum zweiten Initiierungszeitpunkt für Transferaufgabe I ($t(356) = -2.026, p = .044$) und die Überzeugung, Textaufgabe II korrekt gelöst zu haben ($t(231) = -2.497, p = .015$), fort.

Zum zweiten Messzeitpunkt (s. Anhang A18) unterscheidet sich die Akkuratheit der Selbsteinschätzung in Abhängigkeit von der HZB nur noch für die Überzeugung, die Transferaufgabe

richtig gelöst zu haben ($t(63) = -2.302, p = .025$). Auch hier schätzen sich die Studierenden, die eine allgemeinbildendes Gymnasium besuchten, akkurater ein.

Motivation

Unterschiede in der Akkuratheit der Selbsteinschätzung in Abhängigkeit von der angegebenen Motivation liegen insbesondere für den ersten Messzeitpunkt vor (s. Anhang A19). Zum ersten Initiierungszeitpunkt der Selbsteinschätzung ist die Selbsteinschätzung bezüglich des Abschneidens bei der Text- ($t(364) = -4.099, p < .001$) und der Transferaufgabe ($t(364) = -3.421, p = .001$) für Studierende, die angaben, wenig motiviert zum Lernen der Inhalte des externen Rechnungswesens zu sein, akkurater als für Studierende mit hoher Motivation. Diese Tendenz setzt sich am folgenden Initiierungszeitpunkt fort.

Am zweiten Messzeitpunkt haben sich die Unterschiede weitgehend aufgehoben (s. Anhang A20). Lediglich die Selbsteinschätzung hinsichtlich der Transferaufgabe ist an Initiierungszeitpunkt I ($t(152) = -2.531, p = .012$) und Initiierungszeitpunkt II ($t(152) = -3.013, p = .003$) für Studierenden mit geringer Motivation akkurater als für Studierende mit hoher Motivation.

Studierende mit geringer Motivation unterscheiden sich bezogen auf die Performanz bei der Bearbeitung der Aufgaben signifikant von Studierenden mit hoher Motivation. So erreichen Studierende mit höherer Motivation bei der Testung zum ersten Messzeitpunkt insgesamt 12.25 Punkte ($SD = 6.47$) im Mittel, wohingegen Studierende mit geringerer Motivation nur 8.17 Punkte ($SD = 5.61$) erzielen ($t(369) = -3.549, p < .001$). Dieser Zusammenhang zeigt sich für alle Aufgabenformate. Zum zweiten Messzeitpunkt werden die Unterschiede zwischen Studierenden mit höherer und niedrigerer Motivation kleiner, bleiben jedoch für alle Aufgaben mit Ausnahme der Transferaufgabe ($t(152) = .550, p = .583$) signifikant. In der Summe weisen Studierende mit höherer Motivation ($M = 19.91, SD = 4.57$) auch am zweiten Messzeitpunkt signifikant ($t(152) = -3.290, p = .001$) höhere Performanzwerte auf als Studierende mit niedrigerer Motivation ($M = 17.24, SD = 5.50$).

6.3 Zusammenhänge zwischen akkurater Selbsteinschätzung und Performanz

Auf Basis der bisherigen Forschungsergebnisse (s. Kapitel 3) wird angenommen, dass ein positiver Zusammenhang zwischen akkurater Selbsteinschätzung und Performanz existiert.

H₂ Eine akkuratere Selbsteinschätzung steht im wirtschaftswissenschaftlichen Kontext in Zusammenhang mit einem besseren Lernergebnis (Dunlosky & Rawson, 2012; Nietfeld et al., 2006; Thiede et al., 2003).

Zur Prüfung der Hypothese ist es zunächst erforderlich, die Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt bei den Studierenden zu bestimmen, die zu beiden Messzeitpunkten an der Befragung teilgenommen haben, um diese anschließend zur Performanz zum zweiten Messzeitpunkt in Beziehung zu setzen.

Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)

Die Selbsteinschätzung vor der Bearbeitung der Aufgaben ist zum ersten Messzeitpunkt nur in wenigen Fällen akkurat (s. Abbildung 23). Eine besonders ausgeprägte Überschätzung des eigenen Wissens- und Könnensstandes ist für die Bearbeitung der Textaufgaben (66.1 % der Studierenden) sowie der Buchungsaufgaben (44.6 % der Studierenden) zu beobachten. Zur Unterschätzung neigten Studierende besonders beim Lösen der MC-Aufgaben (51.8 % der Studierenden). Im Vergleich zu den anderen Aufgaben besonders akkurat war die Selbsteinschätzung hinsichtlich der Bearbeitung der Transferaufgabe (55.4 % der Studierenden).

Im Durchschnitt beträgt die betragsmäßige Abweichung zwischen erreichter normierter Punktzahl und angegebener Selbsteinschätzung vor der Aufgabenbearbeitung für die Textaufgaben 1.78 Skalenpunkte (SD = 1.13), für die MC-Aufgaben 1.38 Skalenpunkte (SD = .80), für die Buchungsaufgaben 1.51 Skalenpunkte (SD = 1.11) und für die Transferaufgabe 1.49 Skalenpunkte (SD = 1.12). Zur weiteren Prüfung der Akkuratheit der Selbsteinschätzungen wurden weiterhin die Korrelationen zwischen Selbsteinschätzungen und Performanz für jedes Aufgabenformat berechnet (s. Anhang A21). Sie sind für die Bearbeitung und Selbsteinschätzung der Textaufgaben nicht signifikant ($r = .240$, $p = .074$) und erreichen für die Bearbeitung und Selbsteinschätzung der MC-Aufgaben ($r = .405$, $p = .002$) Korrelationswerte schwacher Stärke. Der stärkste (negative) Zusammenhang ergibt sich für die Einschätzung, noch nicht über das nötige Verständnis für die Grundlagen der Buchführung zu verfügen (SE122) und der Performanz bei Buchungsaufgaben ($r = -.544$, $p < .001$).

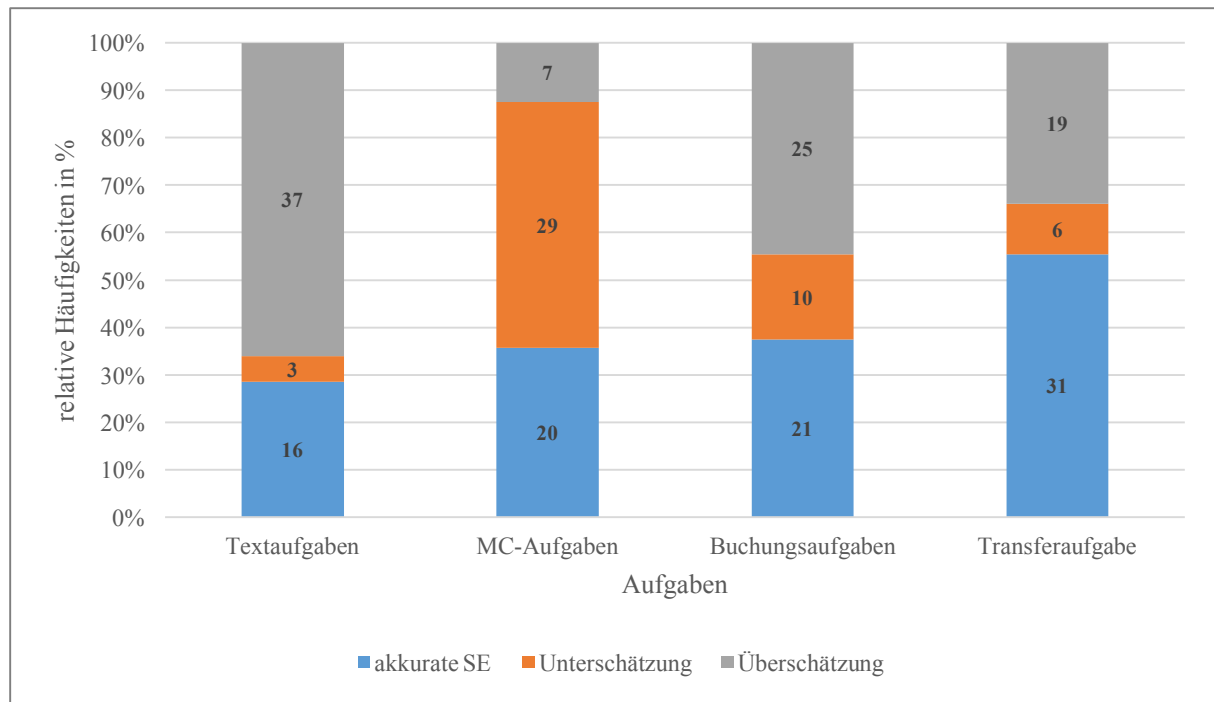


Abbildung 23: Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung an Messzeitpunkt I (N = 56)

Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung (Initiierungszeitpunkt II)

Die Akkuratheit der Selbsteinschätzung nach der Bearbeitung der Aufgaben ist am ersten Messzeitpunkt für die meisten Aufgabenformate höher als vor der Aufgabenbearbeitung (s. Abbildung 24). Die höchste Übereinstimmung zwischen Selbsteinschätzung und Performanz lässt sich erneut für die Bearbeitung der Transferaufgabe I beobachten (62.5 % der Studierenden). Nichtsdestotrotz ist der Anteil an inakkuraten Selbsteinschätzungen auch zu diesem Initiierungszeitpunkt der Selbsteinschätzung in Teilen hoch. Dies gilt insbesondere für die Selbsteinschätzung zum Abschneiden bei MC-Aufgaben (59.0 % der Studierenden unterschätzen sich) und Buchungsaufgaben I (53.6 % der Studierenden unter- bzw. überschätzen sich). Die Betrachtung des Durchschnitts der Abweichungen zwischen erreichter normierter Punktzahl und angegebener Selbsteinschätzung unterstützt die Tendenz akkuraterer Selbsteinschätzungen zum zweiten Initiierungszeitpunkt. So sind die betragsmäßigen Abweichungen für die Textaufgaben ($t(54) = 4.724$, $p < .001$), die Buchungsaufgaben ($t(53) = 3.235$, $p = .002$) und die Transferaufgabe ($t(47) = 2.468$, $p = .017$) signifikant niedriger als noch zum ersten Initiierungszeitpunkt der Selbsteinschätzung. Lediglich die Selbsteinschätzung bezogen auf die Bearbeitung der MC-Aufgaben weist keine signifikante Änderung zum zweiten Initiierungszeitpunkt der Selbsteinschätzung auf ($t(54) = -1.058$, $p = .295$). Auch auf Basis der berechneten Korrelationen zwischen Selbsteinschätzung und Performanz bestätigt sich die Tendenz, dass die Akkuratheit zum späteren Initiierungszeitpunkt gestiegen ist (s. Anhang A22). So sind die Zusammenhänge

nun überwiegend von schwacher ($r = .307$, $p = .021$ für die Bearbeitung und Selbsteinschätzung der Textaufgaben) bis mittlerer ($r = .667$, $p < .001$ für die Bearbeitung und Selbsteinschätzung der Buchungsaufgaben II) Stärke.

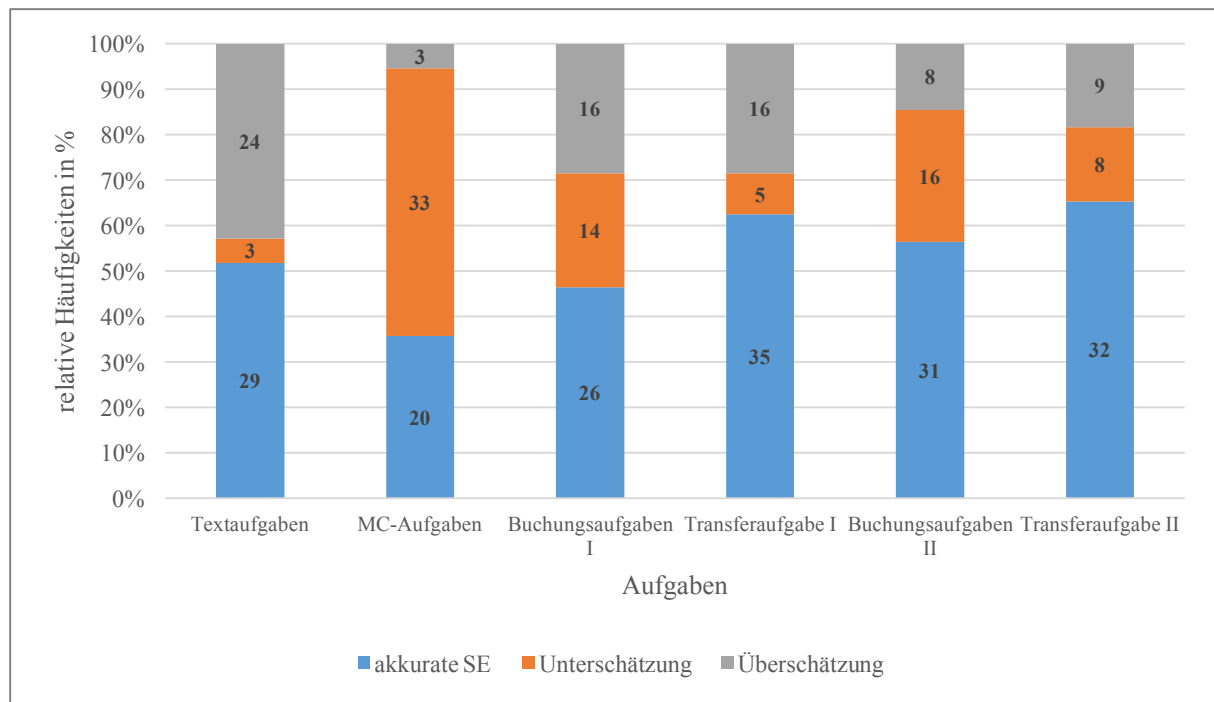


Abbildung 24: Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung an Messzeitpunkt I ($49 \leq n \leq 56$)

Akkuratheit der Überzeugungen

Neben den beiden bislang betrachteten Initiierungszeitpunkten der Selbsteinschätzung wurden die Studierenden nach jeder Aufgabe gebeten, ihre Überzeugung anzugeben, die entsprechenden Aufgaben richtig gelöst zu haben. Die Akkuratheit dieser Selbsteinschätzungen variiert in Abhängigkeit von der betrachteten Aufgabenstellung (s. Abbildung 25). Die höchsten Akkuratheitswerte können für die Selbsteinschätzung zur zweiten Textaufgabe (73.8 % der Studierenden), der Buchungsaufgaben (54.0 % der Studierenden) und der Transferaufgabe (61.5 % der Studierenden) festgestellt werden. Diese Selbsteinschätzungen sind mehrheitlich akkurat. Bei den drei weiteren abgefragten Überzeugungen überwiegen die Fehleinschätzungen. Bei der ersten (42.6 % der Studierenden) und dritten Textaufgabe (39.6 % der Studierenden) liegen öfter Überschätzungen vor, wohingegen bei den MC-Aufgaben Unterschätzungen am häufigsten auftreten (63.6 % der Studierenden). Entsprechend betragen die durchschnittlichen Abweichungen zwischen $M = .67$ Skalenpunkte ($SD = 1.10$) für die zweite Textaufgabe und $M = 1.66$ Skalenpunkte ($SD = 1.33$) für die erste Textaufgabe. Die Korrelationen sind für das Lösen der Buchungsaufgaben ($r = .727$, $p \leq .001$) besonders hoch. Für die Textaufgaben liegen dagegen keine signifikanten Korrelationen vor.

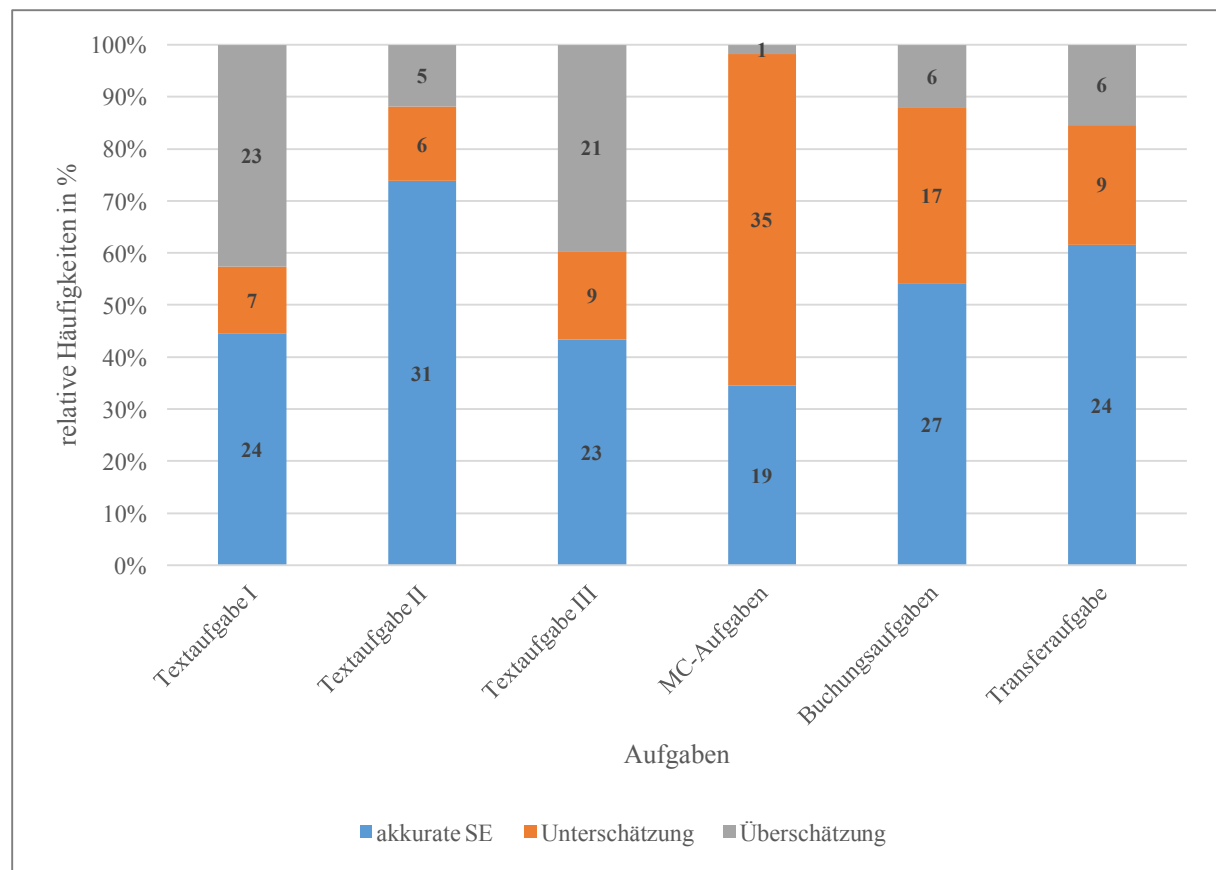


Abbildung 25: Akkuratheit der Überzeugungen an Messzeitpunkt I ($39 \leq n \leq 55$)

Insgesamt zeigt sich also, dass die Akkuratheit der Selbsteinschätzung zu beiden Initiierungszeitpunkten am ersten Messzeitpunkt variiert und in etwa 50 % der Fälle eine Fehleinschätzung vorliegt. Der Anteil ist für die Selbsteinschätzung vor der Aufgabenbearbeitung höher als nach der Aufgabenbearbeitung.

Prüfung des Zusammenhangs von akkurater Selbsteinschätzung und Performanz

Die zweite Hypothese postuliert, dass Studierende mit einer akkurateren Selbsteinschätzung zum ersten Messzeitpunkt am Ende des Semesters (Messzeitpunkt II) eine höhere Performanz aufweisen als Studierende mit weniger akkuraten Selbsteinschätzungen. Zur Prüfung der Hypothese wurde folgendes Verfahren gewählt: Einerseits wurde der Fokus auf die Akkuratheit der Selbsteinschätzungen bezogen auf die **einzelnen Aufgaben** gelegt und diese mit der Performanz bezüglich der einzelnen Aufgaben in Beziehung gesetzt. Andererseits wurde in einem zweiten Schritt die Akkuratheit der Selbsteinschätzungen über **alle Aufgaben** hinweg aggregiert betrachtet und in Bezug zur Performanz gesetzt.

Die zum zweiten Messzeitpunkt gezeigte Performanz in den verschiedenen Aufgaben wurde auf Unterschiede in Abhängigkeit von der Akkuratheit der Selbsteinschätzung bzgl. der verschiedenen Aufgaben vor bzw. nach der Aufgabenbearbeitung am ersten Messzeitpunkt geprüft. Dazu wurden für jede Selbsteinschätzung zwei Gruppen gebildet, wovon eine Studierende mit akkuraten Selbsteinschätzungen und eine mit inakkuraten Selbsteinschätzungen (Über- und Unterschätzung) umfasst. Als Erstes wurde ein t-Test für die Performanz in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) für jedes Aufgabenformat separat durchgeführt. Es zeigt sich, dass sowohl für die Textaufgaben ($t(54) = 1.225$, $p = .226$) und MC-Aufgaben ($t(54) = 1.071$, $p = .289$) als auch für die Buchungsaufgaben ($t(54) = -.872$, $p = .387$) und die Transferaufgabe ($t(54) = 1.901$, $p = .064$) keine signifikanten Mittelwertunterschiede in der Performanz zum zweiten Messzeitpunkt in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung am ersten Messzeitpunkt bestehen. Die Berechnungen wurden für den zweiten Initiierungszeitpunkt der Selbsteinschätzung nach der Aufgabenbearbeitung wiederholt. Erneut konnten bei der Betrachtung der aufgabenspezifischen Selbsteinschätzungen weder für Textaufgaben ($t(54) = 1.523$, $p = .134$) noch für MC-Aufgaben ($t(54) = .416$, $p = .679$) signifikante Mittelwertunterschiede ermittelt werden. Bei den Buchungsaufgaben ergibt sich für die Selbsteinschätzung nach der Bearbeitung der Text- und MC-Aufgaben ein signifikanter Unterschied ($t(54) = 2.330$, $p = .025$) dahingehend, dass die Gruppe mit inakkuratere Selbsteinschätzung höhere Performanzwerte ($M = 10.70$, $SD = 1.47$) aufweist als die Gruppe mit akkurater Selbsteinschätzung ($M = 9.31$, $SD = 2.72$). Bei der wiederholten Selbsteinschätzung nach der Bearbeitung der Buchungsaufgaben liegt jedoch kein signifikanter Mittelwertunterschied mehr vor ($t(54) = 1.514$, $p = .136$). Für die Selbsteinschätzung zur Bearbeitung der Transferaufgaben sind sowohl nach der Bearbeitung der Text- und MC-Aufgaben ($t(54) = .961$, $p = .341$) als auch nach der Bearbeitung der Transferaufgabe keine Mittelwertunterschiede vorhanden ($t(54) = 1.851$, $p = .076$).

Im zweiten Analyseschritt, in dem die Selbsteinschätzungen über alle Aufgaben hinweg aggregiert betrachtet werden, erfolgt die Gruppenbildung indem alle betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz summiert werden. Somit wird für jeden Studierenden ein Gesamtwert ermittelt, der die Abweichungen seiner Selbsteinschätzungen zur Performanz über alle Aufgaben hinweg angibt. Erneut wurde eine Gruppe mit akkurateren Selbsteinschätzungen, die jene Studierende mit einem verhältnismäßig niedrigen Gesamtwert der Abweichungen enthält, und eine mit inakkurateren Selbsteinschätzungen (verhältnismäßig hohe

Gesamtwerte der Abweichungen) gebildet. Für die Gruppenteilung wurde dabei auf den Median der ermittelten Abweichungen zwischen Selbsteinschätzungen und Performanz zurückgegriffen.

Ein t-Test ergibt signifikante Mittelwertunterschiede bezüglich der Performanz ($t(54) = -2.108$, $p = .040$) zwischen beiden Gruppen, wobei die Gruppe der Studierenden mit weniger akkuraten Selbsteinschätzungen ($M = 21.67$ Punkte, $SD = 4.42$) die Gruppe mit akkurateren Selbsteinschätzungen ($M = 18.52$ Punkte, $SD = 6.49$) am ersten Initiierungszeitpunkt in der Performanz zum zweiten Messzeitpunkt übertraf. Diese Differenz resultiert vordergründig aus dem Abschneiden der Studierenden bei den MC-Aufgaben und den Buchungsaufgaben. So schnitten Studierende mit inakkuraterer Selbsteinschätzung bei beiden Aufgabenformaten signifikant besser ab als Studierende mit akkuraterer Selbsteinschätzung (s. Anhang A23).

Betrachtet man die Selbsteinschätzungen nach der Aufgabenbearbeitung (Initiierungszeitpunkt II), liegt hinsichtlich der Gesamtpunktzahl ein signifikanter Mittelwertunterschied vor ($t(55) = -2.903$, $p = .005$). Studierende, die sich zum ersten Messzeitpunkt inakkurater einschätzten, erreichten am zweiten Messzeitpunkt bessere Ergebnisse ($M = 22.00$ Punkte, $SD = 4.37$) als Studierende mit akkuraterer Selbsteinschätzung ($M = 17.86$ Punkte, $SD = 6.27$). Diese Differenzen ergeben sich aus dem besseren Abschneiden der Studierenden mit inakkuraterer Selbsteinschätzung bei den Text-, MC- und Buchungsaufgaben. Keine Unterschiede ergaben sich dagegen bei der Bearbeitung der Transferaufgabe (s. Anhang A23).

Abschließend werden die gleichen oben beschriebenen Analyseschritte für die Überzeugungen vorgenommen. Die Akkuratheitswerte bezüglich der Überzeugungen dienen als Basis zur Bildung zweier Gruppen, von denen eine die akkurateren Überzeugungen umfasst und die andere die weniger akkuraten. Erneut werden die Gruppen zunächst aufgabenspezifisch gebildet, bevor eine aggregierte Betrachtung folgt.

Wie bereits bei den kriterienorientierten Selbsteinschätzungen zum ersten und zweiten Initiierungszeitpunkt zeigen sich bei der aufgabenspezifischen Betrachtung keine Mittelwertunterschiede zwischen Studierenden mit akkurateren und inakkurateren Überzeugungen am ersten Messzeitpunkt hinsichtlich der Performanz zum zweiten Messzeitpunkt. Darüber hinaus wurde die erreichte Gesamtpunktzahl den Überzeugungen gegenübergestellt. Hier zeigt sich ein signifikanter Mittelwertunterschied ($t(55) = -2.462$, $p = .017$) zwischen Studierenden mit akkurateren und weniger akkuraten Überzeugungen hinsichtlich der Richtigkeit der Aufgabenbearbeitungen. Demnach schneiden ähnlich wie bei den kriterienorientierten Selbsteinschätzungen Studierende mit inakkurateren Selbsteinschätzungen am ersten Messzeitpunkt zum zweiten

Messzeitpunkt besser ab ($M = 21.72$ Punkte, $SD = 5.02$) als Studierende mit akkurateren Selbsteinschätzungen am ersten Messzeitpunkt ($M = 18.14$ Punkte, $SD = 5.94$).

Studierende, die sich zum ersten Messzeitpunkt inakkurater einschätzen, erzielen zum zweiten Messzeitpunkt mehr Punkte. In einem nächsten Schritt soll deswegen der Frage nachgegangen werden, ob sich dieses wenig erwartungskonforme Ergebnis sowohl für Über- wie auch Unterschätzer zeigt. Aufgrund der in Teilen geringen Fallzahlen sind varianzanalytische Auswertungsverfahren nur bedingt einsetzbar, deswegen werden die Befunde auf deskriptivem Niveau näher beleuchtet.

In Tabelle 17 sind die durchschnittlichen Performanzwerte zum zweiten Messzeitpunkt in Abhängigkeit von der Akkuratheit der Selbsteinschätzung bezüglich jedes Aufgabenbereiches zum ersten Messzeitpunkt vor der Aufgabenbearbeitung aufgeführt. Tendenziell zeigt sich, dass die Unterschätzung des eigenen Wissens- und Könnensstandes zu besseren Performanzwerten am zweiten Messzeitpunkt führt. Lediglich für die Selbsteinschätzung bezüglich des Abschneidens bei den Textaufgaben, mit denen überwiegend Faktenwissen erfragt wurde, erreichten Studierende mit Überschätzungen ($n = 37$ Studierende, $M = 3.86$ Punkte, $SD = 2.20$) höhere Werte als Studierende mit Unterschätzungen ($n = 3$ Studierende, $M = 3.67$ Punkte, $SD = 1.53$).

Hinsichtlich der Gesamtpunktzahl zeigt sich die Tendenz, dass akkurate Selbsteinschätzungen bezüglich der Bearbeitung von Buchungsaufgaben zu höheren Performanzwerten ($n = 21$ Studierende, $M = 22.0$ Punkte, $SD = 5.71$) als Unterschätzungen ($n = 10$ Studierende, $M = 21.0$ Punkte, $SD = 4.74$) und Überschätzungen ($n = 25$ Studierende, $M = 18.0$ Punkte, $SD = 5.69$) des eigenen Wissens- und Könnensstandes führen. Die höchsten Performanzwerte ($M = 25.00$ Punkte, $SD = 4.15$) am zweiten Messzeitpunkt erreichten jedoch Studierende, die sich hinsichtlich der Transferaufgabe unterschätzten.

Akkuratheit/Performanz		N	Text- aufgaben		MC- Aufgaben		Buchungs- aufgaben		Transfer- aufgabe		Summe	
			M	SD	M	SD	M	SD	M	SD	M	SD
Textaufgaben	A	16	3.00	2.81	4.38	2.06	8.88	2.96	.88	1.63	17.13	7.34
	U	3	3.67	1.53	4.33	2.08	10.00	1.00	1.67	2.08	19.67	2.08
	Ü	37	3.86	2.20	5.78	1.06	10.62	1.75	1.05	1.73	21.32	4.74
MC-Aufgaben	A	20	3.50	2.14	5.00	1.81	9.90	2.22	.80	1.51	19.20	5.50
	U	29	3.76	2.71	5.48	1.53	10.24	2.44	1.45	1.90	20.93	6.44
	Ü	7	3.29	1.38	5.43	1.13	10.00	1.63	.00	.00	18.71	2.50
Buchungsaufgaben	A	21	3.95	2.71	5.67	1.39	10.43	1.33	1.95	1.91	22.00	5.71
	U	10	3.50	1.90	5.80	.79	10.90	1.37	.80	1.69	21.00	4.74
	Ü	25	3.36	2.25	4.80	1.85	9.48	2.95	.36	1.11	18.00	5.69
Transferaufgabe	A	31	3.23	2.36	4.84	1.77	9.39	2.64	.65	1.38	18.10	5.87
	U	6	4.67	2.88	6.33	.82	11.00	1.27	3.00	1.67	25.00	4.15
	Ü	19	3.89	2.16	5.74	1.15	10.95	1.22	1.05	1.81	21.63	4.67

Anmerkung: A = Akkurate Selbsteinschätzung, U = Unterschätzung, Ü = Überschätzung

Tabelle 17: Performanzwerte an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung an Messzeitpunkt I

Die gleichen Berechnungen wurden auch für die Selbsteinschätzung nach der Aufgabenbearbeitung vorgenommen. An dieser Stelle zeigt sich besonders deutlich, dass die Studierenden, die zum ersten Messzeitpunkt zur Unterschätzung neigten, am zweiten Messzeitpunkt höhere Performanzwerte aufweisen als Studierende mit Überschätzung oder akkurater Selbsteinschätzung (s. Tabelle 18). Die höchsten Performanzwerte erzielten dabei erneut jene Studierenden, die sich am ersten Messzeitpunkt bezüglich ihres Abschneidens bei der Transferaufgabe unterschätzten.

Akkuratheit/Performanz			Text- aufgaben		MC- Aufgaben		Buchungs- aufgaben		Transfer- aufgabe		Summe	
			<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>
Textaufgaben	<i>A</i>	29	3.17	2.04	5.03	1.68	9.69	2.75	.76	1.55	18.66	6.11
	<i>U</i>	3	6.00	4.58	5.00	2.65	10.00	1.00	3.00	1.73	24.00	7.21
	<i>Ü</i>	24	3.88	2.25	5.58	1.28	10.50	1.50	.96	1.65	20.92	4.93
MC-Aufgaben	<i>A</i>	20	2.90	1.74	5.15	1.63	9.85	2.23	.30	.98	18.20	4.06
	<i>U</i>	33	4.03	2.62	5.36	1.60	10.15	2.35	1.45	1.87	21.00	6.61
	<i>Ü</i>	3	4.00	1.73	5.00	1.00	10.33	1.16	.00	.00	19.33	3.06
Buchungsaufgaben I	<i>A</i>	26	3.42	2.04	4.88	1.84	9.31	2.72	1.04	1.71	18.65	6.69
	<i>U</i>	14	3.93	2.92	5.79	1.12	10.86	1.35	1.57	1.95	22.14	5.72
	<i>Ü</i>	16	3.69	2.33	5.44	1.32	10.56	1.59	.31	1.01	20.00	3.43
Transferaufgabe I	<i>A</i>	35	3.46	2.39	4.83	1.79	9.66	2.61	.80	1.53	18.74	6.22
	<i>U</i>	5	5.60	1.95	6.20	.84	10.80	1.30	2.80	1.79	25.40	4.51
	<i>Ü</i>	16	3.38	2.13	5.94	.57	10.69	1.20	.75	1.61	20.75	3.87
Buchungsaufgaben II	<i>A</i>	31	3.42	2.26	5.19	1.68	9.94	2.07	1.16	1.75	19.71	5.74
	<i>U</i>	16	3.81	2.76	5.69	1.01	10.56	1.32	1.38	1.89	21.44	4.89
	<i>Ü</i>	8	4.13	1.96	5.38	1.51	10.88	1.34	.00	.00	20.38	3.70
Transferaufgabe II	<i>A</i>	32	3.13	2.17	4.97	1.60	9.66	2.67	.75	1.48	18.50	5.86
	<i>U</i>	8	4.38	2.62	6.38	.74	11.00	1.20	2.75	1.83	24.50	3.67
	<i>Ü</i>	9	4.44	3.25	5.78	.83	10.67	1.00	.89	1.76	21.78	4.60

Anmerkung: *A* = Akkurate Selbsteinschätzung, *U* = Unterschätzung, *Ü* = Überschätzung

Tabelle 18: Performanzwerte an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung an Messzeitpunkt I

Zusammenfassend ist in Abbildung 26 die Gesamtperformanz an beiden Messzeitpunkten in Abhängigkeit von der Akkuratheit der Selbsteinschätzung an Initiierungszeitpunkt I zum ersten Messzeitpunkt für die einzelnen Aufgabenbereiche dargestellt. Es fällt auf, dass die Studierenden zum zweiten Messzeitpunkt durchgängig höhere Performanzwerte erzielten als zum ersten Messzeitpunkt. Allerdings erreichten die Studierenden, die am ersten Messzeitpunkt zur Unterschätzung neigten, bereits an diesem vergleichsweise hohe Punktzahlen in der Testung. Dieser Befund gilt insbesondere für die Studierenden, die ihre Performanz bezüglich kognitiv anspruchsvollerer Aufgaben (Buchungsaufgaben und Transferaufgabe)

unterschätzten. Zum zweiten Messzeitpunkt konnten diese Studierenden ihre Performanzwerte nur unwesentlich steigern. Dahingegen stiegen die Performanzwerte der Studierenden, die sich akkurat einschätzten oder überschätzten zum zweiten Messzeitpunkt deutlich und übertreffen je nach betrachtetem Aufgabenbereich sogar das Niveau der Studierenden, die anfänglich zur Unterschätzung neigten.

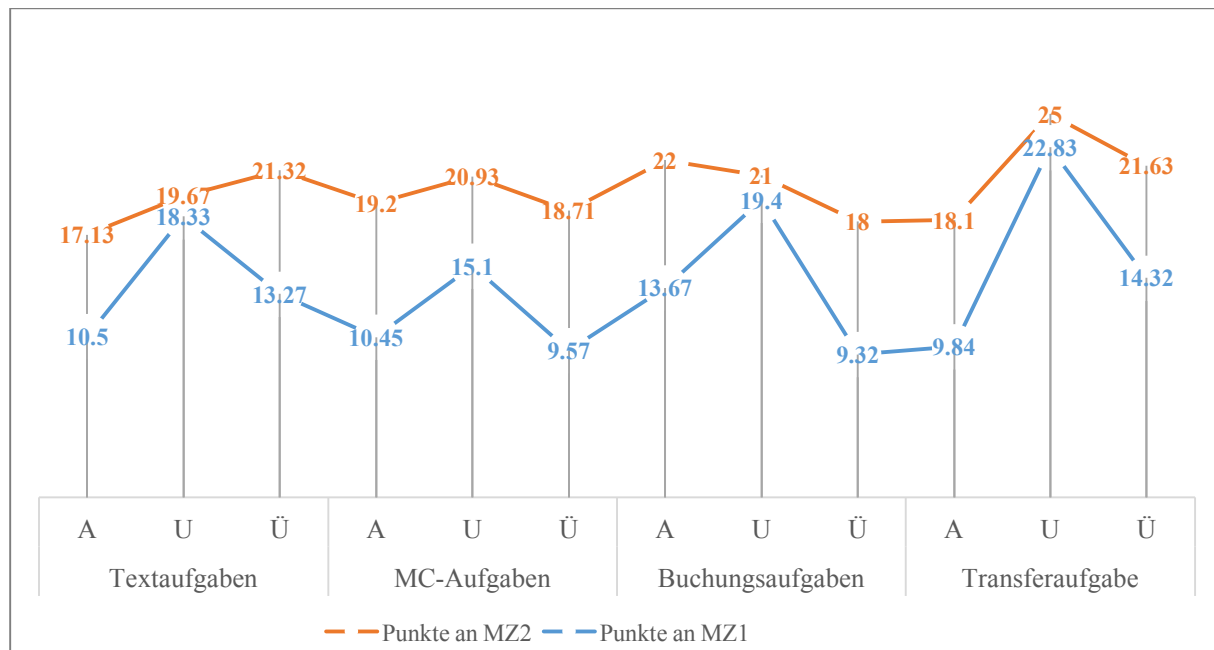


Abbildung 26: Entwicklung der Performanz in Abhängigkeit von der Akkuratheit der Selbsteinschätzung an Messzeitpunkt 1

Ursprüngliches Ziel der Untersuchung war es ferner, zu analysieren, ob es ein quantifizierbares Ausmaß an Über- oder Unterschätzung gibt, welches für die Performanz effektneutral ist.

H₅ Ein gewisses Ausmaß an Fehleinschätzungen ist für die Performanz der Studierenden zum späteren Messzeitpunkt effektneutral (Coutts & Rogers, 1999; Sitzmann & Johnson, 2012).

In dieser Untersuchung hat sich gezeigt, dass eine Fehleinschätzung nicht nur effektneutral, sondern auch im Zusammenhang mit einer günstigeren Entwicklung der Performanz stehen kann. Aufgrund der hohen Stichprobenmortalität vom ersten Messzeitpunkt zum zweiten Messzeitpunkt gestaltet sich die weitere Analyse dieser Hypothese wenig ergiebig. Dennoch besteht in der Betrachtung von Kreuztabellen eine Möglichkeit zur Überprüfung der Annahme. An dieser Stelle soll exemplarisch die Kreuztabelle für die Akkuratheit der Selbsteinschätzung bezogen auf die MC-Aufgaben zum ersten Messzeitpunkt und die jeweilige Performanz zum zweiten Messzeitpunkt gegenübergestellt werden, da zu dieser Aufgabe besonders viele Fehleinschätzungen (Unterschätzung) vorliegen.

		Performanzwerte für MC-Aufgaben zum zweiten Messzeitpunkt							Σ
		1	2	3	4	5	6	7	
Abweichung der Selbsteinschätzung bezogen auf MC-Aufgaben von Performanzwerten zum ersten Messzeitpunkt	-3.00	0	0	0	0	0	1	1	2
	-2.57	0	0	0	0	1	2	0	3
	-2.29	0	0	0	0	0	4	0	4
	-2.14	0	0	1	0	0	0	0	1
	-2.00	0	0	0	0	0	1	0	1
	-1.86	2	0	0	0	1	2	0	5
	-1.57	0	0	0	1	1	0	1	3
	-1.43	0	0	0	0	1	0	0	1
	-1.29	0	0	0	0	0	2	2	4
	-1.14	0	0	0	0	1	1	0	2
	-1.00	0	0	0	0	0	2	1	3
	-0.86	0	0	0	0	0	2	2	4
	-0.57	0	0	0	0	1	1	0	2
	-0.43	0	0	0	0	0	2	1	3
	-0.29	0	0	0	1	1	0	0	2
	0.00	0	0	0	0	0	0	2	2
	0.14	0	0	1	0	0	2	0	3
	0.29	0	0	0	0	0	1	0	1
	0.43	0	0	0	0	0	0	1	1
	0.57	0	0	0	0	0	1	0	1
	0.86	0	1	0	1	0	0	0	2
	1.00	0	0	0	0	1	1	0	2
	1.14	0	0	0	0	1	0	0	1
	1.43	0	1	1	0	0	0	1	3
	1.57	0	1	0	0	0	0	1	2
	1.86	0	0	0	0	0	0	1	1
	2.29	0	0	0	0	1	0	0	1
	3.29	0	0	0	1	0	1	0	2
	Σ	2	3	3	5	9	23	11	56

Tabelle 19: Kreuztabelle für Fehleinschätzungen in Bezug auf MC-Aufgaben zu Messzeitpunkt I und Performanz bei MC-Aufgaben zu Messzeitpunkt II (N = 56 Studierende)

In der Kreuztabelle (s. Tabelle 19) sind die Häufigkeiten der jeweiligen Abweichungen der Selbsteinschätzung vor der Bearbeitung der MC-Aufgaben (Initiierungszeitpunkt I) zum ersten Messzeitpunkt und die zugehörige Performanz bei den MC-Aufgaben zum zweiten Messzeitpunkt abgebildet. Mit verschiedenen Farbtönen wird die Häufigkeit der Zuordnungen veranschaulicht, wobei orange Farbtöne für Unterschätzung, blaue Farbtöne für akkurate

Selbsteinschätzungen und graue Farbtöne für Überschätzungen stehen. Die Betrachtung dieser Kreuztabelle sowie auch jener für die weiteren Selbsteinschätzungen, die an dieser Stelle nicht aufgeführt sind, zeigt kein erkennbares Muster dahingehend, dass ein gewisses Ausmaß der Fehleinschätzung zum ersten Messzeitpunkt in Zusammenhang mit höheren Performanzwerten zum zweiten Messzeitpunkt stehen könnte. Besonders stark ausgeprägte Unterschätzungen zum ersten Messzeitpunkt stehen im Zusammenhang mit guten Testergebnissen am zweiten Messzeitpunkt. Es zeigt sich ferner erneut, dass eine Unterschätzung zum ersten Messzeitpunkt tendenziell eher mit höheren Performanzwerten zum zweiten Messzeitpunkt zusammenhängt als eine Überschätzung.

6.4 Lernstrategien und Lernzeit in Abhängigkeit von der Akkuratheit der Selbsteinschätzung

- H₃ Studierende, die zur Unterschätzung neigen, setzen mehr Strategien zur Steuerung des weiteren Lernprozesses (Lernstrategien) ein als Studierende mit akkurater Selbsteinschätzung (Thiede et al., 2003).
- H₄ Zur Unterschätzung tendierende Studierende verbringen mehr Zeit mit dem Lernen der Inhalte als Studierende mit akkurater Selbsteinschätzung (Dunlosky & Rawson, 2012).

Es wurde angenommen, dass sich der Einsatz von Lernstrategien und die aufgebrauchte Lernzeit in Abhängigkeit von der Akkuratheit der Selbsteinschätzung unterscheiden. Dies kann ein Grund für die unterschiedlichen Performanzwerte in Abhängigkeit von der Akkuratheit der Selbsteinschätzung sein (s. Kapitel 6.2 und 6.3).

Die von Studierenden eingesetzten Lernstrategien wurden anhand von 18 Items erfasst, die sich in Teilen an bestehende Instrumente orientierten (s. Kapitel 5.2.2.3). In Anhang A24 sind Mittelwerte und Standardabweichungen der Gesamtstichprobe für alle 18 Items zu den Lernstrategien sowie für die vier Items zum Zeitaufwand für beide Messzeitpunkte dargestellt. Der Veranstaltungsbesuch geht zum zweiten Messzeitpunkt zurück, was jedoch daran liegen könnte, dass am Semesterende eher eine Veranstaltung versäumt wurde als zur Semestermitte⁴². Hinsichtlich der Vor- und Nachbereitung der Veranstaltungen sowie des Einsatzes der weiteren Lernstrategien zeigen sich auf deskriptiver Ebene zum zweiten Messzeitpunkt dagegen höhere Mittelwerte. Besonders zugestimmt wurde den Items zur Nachbereitung von Vorlesung ($M = 4.20$, $SD = 1.75$) und Tutorium ($M = 4.29$, $SD = 1.71$) sowie der Nutzung der vorlesungsbegleitenden Literatur ($M = 4.17$, $SD = 1.75$). Die Studierenden gaben zum zweiten Messzeitpunkt an, mehr Zeit aufzuwenden als noch zum ersten Messzeitpunkt. Der Aufwand ist aber dennoch nicht höher als im Vergleich zu anderen Veranstaltungen des Studiums ($M = 3.30$, $SD = 1.24$).

Vor der konkreten Prüfung der Hypothesen anhand der Teilstichprobe, die zu beiden Messzeitpunkten an der Befragung teilnahm, werden zunächst die Ergebnisse für die gesamte Stichprobe auf Skalenebene angeführt.

⁴² Die Items zum Veranstaltungsbesuch wurden derart codiert, dass ein einmaliges Fehlen im Semester bereits zu einer Abstufung führte (s. Anhang A2).

Abhängig von der Akkuratheit der Selbsteinschätzung⁴³ zu den *Textaufgaben*, den *Buchungsaufgaben* und der *Transferaufgabe* vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) ergeben sich am ersten Messzeitpunkt signifikante Mittelwertunterschiede in der Lernstrategienutzung (s. Anhang A25). Demnach geben Studierende mit inakkuraten Selbsteinschätzungen an, häufiger die *Veranstaltung vor- und nachzubereiten* als Studierende mit akkuraterer Selbsteinschätzung. Außerdem stimmen sie den Items zur *Aufgabenbearbeitung* signifikant stärker zu. Insgesamt betrachtet sind die Mittelwerte jedoch für beide Gruppen relativ niedrig und erreichen die Skalenmitte von 3.5 nur für die Subskala *Umgang mit Schwierigkeiten*.

Vergleichbare Befunde ergeben sich für die Betrachtung der Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung (Initiierungszeitpunkt II). Differenzen in der angegebenen Realisierung der Lernstrategien zeigen sich vor allem hinsichtlich der *Vor- und Nachbereitung der Veranstaltungen* und der *Bearbeitung der Aufgaben*. Ähnlich wie zum ersten Initiierungszeitpunkt sind diese Items stärker bei jenen Studierenden ausgeprägt, die ihren Wissens- und Könnensstand inakkurater einschätzten (s. Anhang A26).

Am zweiten Messzeitpunkt liegen die Mittelwerte allgemein höher als beim ersten Messzeitpunkt und übersteigen häufiger die Skalenmitte. Signifikante Unterschiede zwischen den realisierten Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen liegen nun kaum noch vor (s. Anhang A27 und Anhang A28).

Zur Prüfung der Hypothese 3 wurde die Akkuratheit der Selbsteinschätzung für die Teilstichprobe, die Studierende umfasst, die an beiden Messzeitpunkten an der Befragung teilgenommen hat, zum ersten Messzeitpunkt mit den angegebenen realisierten Lernstrategien zum zweiten Messzeitpunkt verglichen.

In Tabelle 20 sind die Mittelwerte der Items zu den eingesetzten Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzung dargestellt. Hierzu wurden die Selbsteinschätzungen der Studierenden erneut über das zweite Akkuratheitsmaß der Selbsteinschätzungen (betragsmäßige Abweichungen) in eine akkuratere und eine inakkuratere Gruppe zusammengefasst. Außerdem wird die Akkuratheit auf die verschiedenen Aufgabenformate bezogen angezeigt. Aus Gründen der Übersichtlichkeit enthält Tabelle 20 nur die signifikanten Ergebnisse der t-Tests.

⁴³ Zur Bestimmung der Akkuratheit der Selbsteinschätzung wurde das zweite Akkuratheitsmaß, also die betragsmäßige Abweichung zwischen Selbsteinschätzung und Performanz, herangezogen (vgl. Kapitel 6.2).

Auffällig ist, dass Studierende mit inakkurateren Selbsteinschätzungen angeben, durchgängig mehr Lernstrategien einzusetzen als Studierende mit akkurateren Selbsteinschätzungen. So stimmen sie bspw. den Aussagen, die *Übung vorzubereiten (LS5)* und *bereitgestellte Übungsaufgaben eigenständig zu lösen (LS10)* signifikant stärker zu als Studierende mit akkurateren Selbsteinschätzungen. Diese Tendenz liegt für beide Initiierungszeitpunkte der Selbsteinschätzung vor, wobei die Unterschiede zwischen Studierenden mit akkuraterer und inakkuraterer Selbsteinschätzung zum zweiten Initiierungszeitpunkt hin zunehmen.

			akkurate SE		inakkurate SE				
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
Initiierungszeitpunkt I	TA	LS10	3.29	1.56	4.23	1.45	-2.300	52	.025
		LS14	3.04	1.43	3.86	1.27	-2.255	53	.028
	MC		keine signifikanten Mittelwertunterschiede						
	BS	LS16	3.41	1.76	4.52	1.53	-2.526	54	.015
		LS17	3.59	1.93	4.90	1.32	-2.934	54	.005
	FS		keine signifikanten Mittelwertunterschiede						
Initiierungszeitpunkt II	TA	LS7	3.96	1.40	4.82	1.42	-2.278	54	.027
	MC	LS5	2.21	1.26	3.07	1.73	-2.152	54	.036
	BSI	LS5	2.18	1.31	3.07	1.68	-2.223	54	.030
		LS6	2.75	1.58	3.70	1.64	-2.200	53	.032
		LS10	3.22	1.60	4.33	1.30	-2.798	52	.007
		LS16	3.50	1.73	4.46	1.60	-2.165	54	.035
	FSI	LS10	3.16	1.57	4.11	1.45	-2.246	52	.029
	BSII		keine signifikanten Mittelwertunterschiede						
	FSII	LS5	2.00	1.27	2.86	1.48	-2.129	47	.039
		LS10	3.00	1.41	4.15	1.51	-2.645	45	.011

Anmerkung: TA = Textaufgaben, MC = Multiple Choice-Aufgaben, BS = Buchungsaufgaben, FS = Transferaufgabe

Tabelle 20: Angabe realisierter Lernstrategien zu Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung zu Messzeitpunkt I

In einer weiteren Analyse werden die von den Studierenden eingeschätzten Items zu den eingesetzten Lernstrategien abhängig davon betrachtet, ob eine akkurate Selbsteinschätzung, Unterschätzung oder Überschätzung vorliegt. Hierbei zeigt sich, dass insbesondere Studierende,

die zur Unterschätzung neigen, angaben, Lernstrategien häufiger einzusetzen. In Abbildung 27 sind die Ausprägungen der Lernstrategien zum zweiten Messzeitpunkt beispielhaft in Abhängigkeit von der Akkuratheit der Selbsteinschätzung (Initiierungszeitpunkt I) bezogen auf die Buchungsaufgaben dargestellt. Bei nahezu allen Items sind die Werte für die Gruppe der Unterschätzer am höchsten. Besonders hoch sind die Zustimmungswerte für die Items LS7 bis LS10 sowie LS16⁴⁴. Demnach geben die Studierenden an, häufiger die *Veranstaltungen nachzubereiten* und bei *Schwierigkeiten Kommilitonen um Rat zu bitten*.

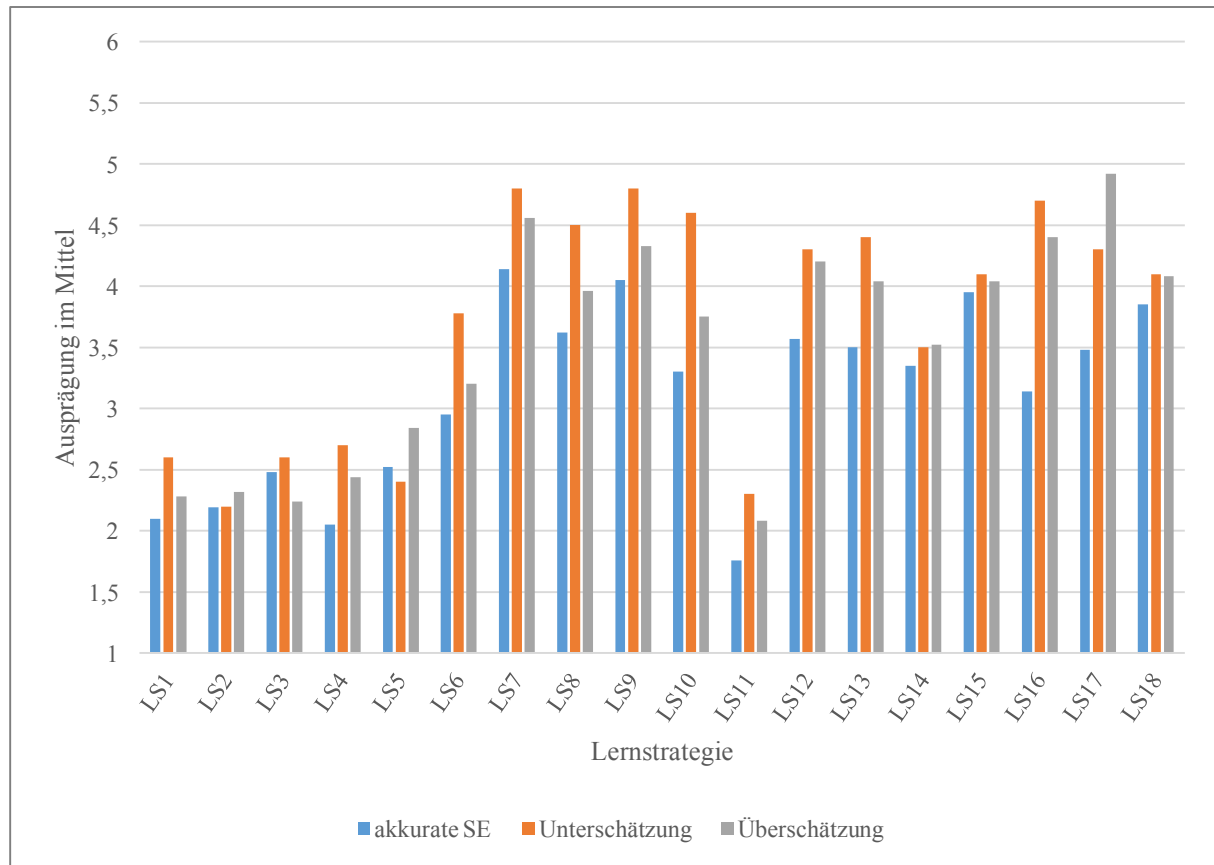


Abbildung 27: Ausprägung der Lernstrategien zu Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Buchungsaufgaben) zu Messzeitpunkt I ($10 \leq n \leq 25$)

Neben den eingesetzten Lernstrategien werden die Unterschiede in der eingesetzten Lernzeit in Abhängigkeit von der Akkuratheit der Selbsteinschätzung bei den verschiedenen Aufgaben untersucht. Hierbei wird zunächst erneut die Gesamtstichprobe zu beiden Messzeitpunkten betrachtet, bevor die Analysen für diejenigen Studierenden, die an beiden Messzeitpunkten an der Befragung teilgenommen haben, dargestellt werden.

⁴⁴ Aus Gründen der besseren Lesbarkeit weichen die Itembezeichnungen in Abbildung 27 von jenen im Codebuch (s. Anhang 2) ab. Die neuen Itembezeichnungen lauten wie folgt: LS2101 → LS1; LS2102 → LS2, etc.

Studierende, die sich vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) hinsichtlich ihrer Performanz bei der Aufstellung von Buchungssätzen inakkurater einschätzen, unterscheiden sich im angegebenen Zeitaufwand für das Lernen der Vorlesungsinhalte im externen Rechnungswesen signifikant von den Studierenden mit akkurateren Selbsteinschätzungen ($t(350) = -2.108, p = .036$). Sie wenden wöchentlich etwa 15 Minuten mehr Zeit auf ($M = 60.09$ Minuten, $SD = 70.57$) als Studierende mit akkurateren Selbsteinschätzungen ($M = 46.04$ Minuten, $SD = 70.57$). Dieses Ergebnis bestätigt sich bei der Betrachtung der Selbsteinschätzungen zur Bearbeitung von Transferaufgaben ($t(350) = -2.423, p = .016$). Erneut liegen die angegebenen Zeitaufwendungen bei Studierenden mit inakkurater Selbsteinschätzung ($M = 61.39$ Minuten, $SD = 70.63$) höher als bei Studierenden mit akkurateren Selbsteinschätzungen ($M = 45.27$ Minuten, $SD = 53.55$). Zusätzlich unterscheiden sich die Studierenden an dieser Stelle auch bei der Einschätzung ihres relativen Zeitaufwandes ($t(363) = -2.748, p = .006$). Studierende mit inakkurateren Selbsteinschätzungen stimmten der Aussage, im Vergleich zu anderen Veranstaltungen des Semesters mehr für das Lernen der Inhalte des externen Rechnungswesens aufgebracht zu haben ($M = 2.57, SD = 1.19$) eher zu als Studierende mit akkurateren Selbsteinschätzungen ($M = 2.22, SD = 1.23$).

Für den zweiten Initiierungszeitpunkt der Selbsteinschätzung nach der Aufgabenbearbeitung bestätigen sich die Unterschiede im angegebenen Zeitaufwand in Abhängigkeit von der Akkuratheit der Selbsteinschätzung. Studierende mit inakkurater Selbsteinschätzung bei der Aufstellung von Buchungssätzen geben an, einen höheren relativen Zeitaufwand gehabt zu haben als Studierende mit akkurateren Selbsteinschätzungen ($t(360) = -1.963, p = .050$). Außerdem schätzen Studierende mit inakkurater Selbsteinschätzung nach der Bearbeitung von Transferaufgaben ihren Zeitaufwand in allen Bereichen signifikant höher ein.

Die gleichen Berechnungen wurden nochmals für den zweiten Messzeitpunkt durchgeführt. Hier zeigen sich für den ersten Initiierungszeitpunkt keine signifikanten Unterschiede im investierten Zeitaufwand in Abhängigkeit von der Akkuratheit der Selbsteinschätzung. Lediglich für den angegebenen Zeitaufwand zur Bearbeitung von Buchungsaufgaben liegen annähernd signifikante Resultate vor ($t(146) = 1.877, p = .063$). Demnach haben Studierende mit akkuraterer Selbsteinschätzung bei den Buchungsaufgaben wöchentlich mehr Zeit für das Lernen der Vorlesungsinhalte aufgebracht ($M = 121.23$ Minuten, $SD = 160.75$) als Studierende mit inakkuraterer Selbsteinschätzung ($M = 79.86$ Minuten, $SD = 96.89$).

Auch für den zweiten Initiierungszeitpunkt konnten am zweiten Messzeitpunkt keine signifikanten Unterschiede im angegebenen Zeitaufwand in Abhängigkeit von der Akkuratheit der Selbsteinschätzung ermittelt werden.

In einem letzten Schritt wird untersucht, ob die Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt in einem Zusammenhang mit dem angegebenen Zeitaufwand zum zweiten Messzeitpunkt steht. Signifikante Unterschiede in der Lernzeit am zweiten Messzeitpunkt bestehen abhängig von der Akkuratheit der Selbsteinschätzung der MC-Aufgaben vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) für die Inhalte der Vorlesung ($t(53) = -2.501$, $p = .016$) und die Inhalte der Tutorien ($t(53) = -2.332$, $p = .024$) dahingehend, dass Studierende mit inakkuraterer Selbsteinschätzung angaben, mehr Zeit aufgewandt zu haben ($M = 133.50$ Minuten, $SD = 158.14$ für die Vorlesungsinhalte und $M = 90.33$ Minuten, $SD = 63.82$ für die Tutoriumsinhalte) als Studierende mit akkuraterer Selbsteinschätzung ($M = 53.40$ Minuten, $SD = 69.29$ für die Vorlesungsinhalte und $M = 53.16$ Minuten, $SD = 52.25$).

Die aufgewandte Lernzeit zum zweiten Messzeitpunkt ergab keine signifikanten Unterschiede in Abhängigkeit von der Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung (Initiierungszeitpunkt II) zum ersten Messzeitpunkt. Tendenziell liegen die Zeitaufwendungen für die Gruppe Studierender mit inakkurateren Selbsteinschätzungen jedoch etwas höher als für jene mit akkurateren Selbsteinschätzungen.

Analog zur Vorgehensweise in Kapitel 6.3 werden in Tabelle 21 die von den Studierenden am zweiten Messzeitpunkt angegebenen Zeitaufwendungen dargestellt, abhängig davon, ob eine akkurate Selbsteinschätzung, Unterschätzung oder Überschätzung zum ersten Initiierungszeitpunkt vorliegt. Hierbei zeigt sich, dass die Studierenden, die zur Unterschätzung neigen, für alle Veranstaltungsformen angaben, mehr Zeit für das Lernen der Inhalte des externen Rechnungswesens zu investieren als Studierende mit akkurater Selbsteinschätzung. Insgesamt übertraf die Gruppe der Studierenden, die sich unterschätzte, mit durchschnittlich 302.77 Minuten Lernzeit die Gruppe der Studierenden mit akkurater Selbsteinschätzung ($M = 208.94$ Minuten) deutlich. Auch die Studierenden, die ihren eigenen Wissens- und Könnensstand eher überschätzten, weisen höhere Zeitaufwendungen auf ($M = 217.2$ Minuten) als die Studierenden mit akkurater Selbsteinschätzung, jedoch weniger deutlich als jene, die sich zum ersten Messzeitpunkt unterschätzten.

	Textaufgabe	MC-Aufgaben	Buchungsaufgaben	Transferaufgabe	Ø
	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	
Akkurate SE	214.31	141.89	219.2	260.35	208.94
Vorlesung	98	52.63	84.1	111.77	86.5
Übung	51.56	43	63.65	69.26	56.87
Tutorien	64.75	46.26	71.45	79.32	65.44
Unterschätzung	401.66	296.72	268.5	244.17	302.77
Vorlesung	193.33	126.21	90	108.33	129.47
Übung	103.33	80.17	76.5	64.17	81.04
Tutorien	105	90.34	102	71.67	92.25
Überschätzung	224.77	225.71	232.48	185.84	217.2
Vorlesung	86.44	97.14	110.32	68.06	90.49
Übung	65.75	51.43	58.56	53.89	57.41
Tutorien	72.58	77.14	63.6	63.89	69.30

Tabelle 21: Zeitaufwand an Messzeitpunkt II in Abhängigkeit von der Akkuratheit der Selbsteinschätzung zu Initiierungszeitpunkt I an Messzeitpunkt I

7 Diskussion und Ausblick

7.1 Zusammenfassung zentraler Ergebnisse

Da im Rahmen dieser Untersuchung eine Vielzahl von Maßen zur Akkuratheit der Selbsteinschätzung und ihrer Zusammenhänge mit der Performanz bestimmt wurden, sollen die Ergebnisse vor der Diskussion der Befunde in Kapitel 7.2 überblicksartig zusammengefasst werden.

Akkuratheit der Selbsteinschätzung

Erste Aufschlüsse zur Akkuratheit der Selbsteinschätzung lieferte die Betrachtung der von den Studierenden zu Beginn der Befragung vorgenommenen Einstufung ihres Wissens- und Könnensstandes anhand der verschiedenen Taxonomiestufen (SE121 bzw. SE221). Hierbei zeigte sich, dass die Studierenden, die angaben, noch nicht über das nötige Verständnis über die Grundlagen der Buchführung zu verfügen, auch die schlechtesten Performanzwerte in der anschließenden Testung aufwiesen, während Studierende, die bereits glaubten, Transferaufgaben lösen zu können, insgesamt die besten Performanzwerte zeigten. Allerdings bedeutete dies nicht, dass sie auch tatsächlich die Transferaufgabe lösen konnten.

Zur genaueren Bestimmung der Akkuratheit der Selbsteinschätzung wurden in der Folge drei Maße berechnet. Zum einen erfolgte auf Basis der **absoluten Abweichungen (AA)** zwischen Performanz und Selbsteinschätzung eine Einteilung nach akkurater Selbsteinschätzung, Unter- oder Überschätzung. Eine Fehleinschätzung lag demnach dann vor, wenn die Selbsteinschätzung mindestens einen Skalenpunkt von der tatsächlichen Performanz abwich (vgl. Vorgehensweise bei Coutts & Rogers, 1999). Ein zweites Maß gibt die **betragsmäßigen Abweichungen (BA)** zwischen Selbsteinschätzung und Performanz an. Je höher dieser Wert ist, umso größer ist die Fehleinschätzung. Allerdings erlaubt dieses Maß keine Rückschlüsse darauf, ob eine Über- oder Unterschätzung vorliegt. Zuletzt wurden als drittes Maß die **Korrelationen nach Pearson** zwischen Selbsteinschätzung und Performanz ermittelt. Die Korrelationen wurden dabei zum einen für die Zusammenhänge zwischen den auf der Likert-Skala vorgenommenen Selbsteinschätzungen und der Performanz berechnet (K_L). Zum anderen wurde auch die anfängliche Selbsteinschätzung, die eine Einstufung des eigenen Wissens und Könnens auf eine Taxonomieebene vorsah, mit der späteren Performanz korreliert (K_T). Je höher die Korrelationswerte in beiden Fällen sind, umso akkurater sollten die Selbsteinschätzungen der Studierenden sein.

In Tabelle 22 sind die Ergebnisse zur Akkuratheit der Selbsteinschätzung für alle Initiierungs- und Messzeitpunkte dargestellt. Dabei wird auf alle drei Maße Bezug genommen. Das erste Maß, das Aussagen darüber erlaubt, inwiefern eine Unter- oder Überschätzung vorliegt, ist in der ersten Zeile für jeden Initiierungszeitpunkt hinterlegt (AA). Die Farben signalisieren dabei, inwiefern bezogen auf jede Aufgabe eine Überschätzung, eine Unterschätzung oder eine akkurate Selbsteinschätzung vorlag. Die Zuordnung zu einer dieser drei Ausprägungen der Selbsteinschätzung erfolgte dabei nur, wenn mehr als 50 % der Studierenden diese für die jeweilige Aufgabe aufwiesen. Sollte keine der drei Ausprägungen der Selbsteinschätzung dominant sein (gelb), wurde die tendenzielle Ausprägung⁴⁵ angegeben.

In der zweiten Zeile (BA) ist für jeden Initiierungs- und Messzeitpunkt die betragsmäßige Abweichung zwischen Selbsteinschätzung und tatsächlich gezeigter Performanz angegeben, während in der dritten und vierten Zeile die relevanten Korrelationen (K_L und K_T zur Bestimmung der Akkuratheit der Selbsteinschätzung) zusammenfassend dargestellt sind. Die Ergebnisse beziehen sich dabei jeweils auf die Gesamtstichprobe, also für alle Studierende, die an Messzeitpunkt I und/oder an Messzeitpunkt II teilgenommen haben.

⁴⁵ A = akkurate Selbsteinschätzung, U = Unterschätzung, Ü = Überschätzung

Messzeitpunkt I								Messzeitpunkt II									
		TA	MC		BS		FS			TA	MC		BS		FS		
IZP I	AA	Ü	(A)		(Ü)		A	Ü		Ü	(U)		A		Ü		
	BA	1.82 (1.19)		1.30 (0.94)		1.55 (1.18)		1.67 (1.31)		1.80 (1.05)		1.21 (.85)		1.04 (.95)		2.41 (1.20)	
	K _L	.195**		.371**		.496**		.218**		.228**		.246**		.141 ^{n.s.}		.131 ^{n.s.}	
	K _T	.161**		.364**		.441**		.140**		.122 ^{n.s.}		.199*		.324**		.198*	
IZP II	AA	(A)		(U)		I II	A A	I II	A A	Ü		U		I II	A A	I II	Ü Ü
	BA	1.21 (.99)		1.47 (1.10)		I II	1.15 (.97) .92 (.85)	I II	1.32 (1.25) 1.10 (1.17)	1.33 (.95)		1.47 (.96)		I II	1.02 (.82) .84 (.67)	I II	2.35 (1.23) 2.03 (1.20)
	K _L	.269**		.307**		I II	.564** .667**	I II	.254** .303**	.297**		.253**		I II	.438** .568**	I II	.105 ^{n.s.} .307**
Überzeugungen	AA	I II III	A A (A)	U		A			A		I II III	(A) A A	U		A		
	BA	I II III	1.35 (1.33) .49 (.99) 1.44 (1.06)		1.32 (1.03)		1.04 (.94)		.94 (1.30)		I II III	1.70 (1.37) .88 (1.07) 1.30 (.97)		1.17 (.87)		.74 (.70) 1.61 (1.58)	
	K	I II III	.198** .471** .242**		.554**		.601**		.408**		I II III	.308** .696** .036 ^{n.s.}		.416**		.557** .404**	

Tabelle 22: Übersicht zur Akkuratheit der Selbsteinschätzung für die eingesetzten Aufgabentypen zu den verschiedenen Mess- und Initiierungszeitpunkten

Die Betrachtung des ersten Maßes zur Bestimmung der Akkuratheit der Selbsteinschätzung zeigt, dass diese sich zum zweiten Messzeitpunkt verglichen mit dem ersten Messzeitpunkt eher verschlechterte. Sind die Akkuratheitswerte für die Überzeugungen, die Aufgaben korrekt gelöst zu haben für beide Messzeitpunkte auf ähnlichem Niveau, nehmen sie insbesondere für den zweiten Initiierungszeitpunkt zum zweiten Messzeitpunkt ab. So neigten die Studierenden in der Einschätzung, ihr Wissen bereits auf neue Aufgabenstellungen transferieren zu können, zum zweiten Messzeitpunkt zur Überschätzung, wohingegen sie zum ersten Messzeitpunkt noch überwiegend akkurate Selbsteinschätzungen aufwiesen. Allgemein ist die Überschätzung zum zweiten Messzeitpunkt stärker ausgeprägt als zum ersten Messzeitpunkt.

Die betragsmäßigen Abweichungen zwischen Selbsteinschätzung und Performanz (zweites Maß der Akkuratheit der Selbsteinschätzung) unterschieden sich am ersten Initiierungszeitpunkt für Text- und MC-Aufgaben für beide Messzeitpunkte kaum. Bezogen auf die Aufstellung von Buchungssätzen nahmen sie zum zweiten Messzeitpunkt ab, während sie für das Lösen von Transferaufgaben zunahmen. Auch für den zweiten Initiierungszeitpunkt und die Überzeugungen bleiben die Ergebnisse für dieses Maß zur Akkuratheit der Selbsteinschätzung ambivalent. Allerdings zeigt sich, dass das Abschneiden bei der Bearbeitung der Transferaufgabe zu Messzeitpunkt II deutlich inakkurater eingeschätzt wird.

Die Korrelationswerte zum Zusammenhang von Selbsteinschätzung und Performanz (drittes Maß der Akkuratheit der Selbsteinschätzung) sinken für fast alle Aufgaben und Initiierungszeitpunkte von Messzeitpunkt I zu Messzeitpunkt II, was für leicht inakkurater Selbsteinschätzungen am zweiten Messzeitpunkt spricht.

Performanz

Darüber hinaus konnte festgestellt werden, dass die Studierenden am zweiten Messzeitpunkt bei der Bearbeitung der Testaufgaben deutlich besser abschnitten. Insbesondere die Buchungs- und MC-Aufgaben konnten am Semesterende oftmals korrekt gelöst werden, während sich die gezeigte Performanz bezüglich der Text- und Transferaufgabe zwar verbesserte, aber noch immer auf einem recht niedrigen Niveau verharrte.

Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung und Performanz

Der Zusammenhang zwischen akkurater Selbsteinschätzung zum ersten Messzeitpunkt und Performanz zum zweiten Messzeitpunkt wurde für verschiedene Aufgabenstellungen mit verschiedenen Anspruchsniveaus analysiert. Dabei wurden die potentiellen Zusammenhänge sowohl aufgabenspezifisch als auch allgemein über alle Aufgaben hinweg ermittelt.

Zusammenfassend lässt sich festhalten, dass die aufgabenspezifische Betrachtung der Akkuratheit der Selbsteinschätzung oftmals keine Zusammenhänge mit der Performanz aufwies. Wurde dagegen das Ausmaß der Fehleinschätzung über alle Aufgaben summiert und daraufhin eine Einteilung von Studierenden mit akkurateren und inakkurateren Selbsteinschätzungen vorgenommen, konnten insbesondere für den zweiten Initiierungszeitpunkt der Selbsteinschätzung signifikante Zusammenhänge mit der Performanz identifiziert werden. Dabei zeigte sich, dass Studierende mit inakkuraterer Selbsteinschätzung am ersten Messzeitpunkt zum zweiten Messzeitpunkt bessere Performanzwerte aufwiesen als Studierende mit akkuraterer Selbsteinschätzung.

Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung und Lernstrategieinsatz sowie Lernzeit

Hinsichtlich der Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der eingesetzten Lernstrategien sowie der aufgewandten Lernzeit hat sich gezeigt, dass die Studierenden mit inakkurateren Selbsteinschätzungen angaben, häufiger Lernstrategien einzusetzen und mehr Zeit für das Lernen der Inhalte der Veranstaltung aufzuwenden.

Einfluss der Kontrollvariablen auf die Akkuratheit der Selbsteinschätzung

Es wurde geprüft, inwiefern sich die Akkuratheit der Selbsteinschätzung hinsichtlich des Geschlechts, des Typs der Hochschulzugangsberechtigung sowie der Motivation unterschieden. Für das Geschlecht konnten zu allen Initiierungs- und Messzeitpunkten nur wenige signifikante Unterschiede festgestellt werden, weshalb davon ausgegangen wird, dass das Geschlecht keinen Einfluss auf die Akkuratheit der Selbsteinschätzung hat.

Ähnliches gilt für die Hochschulzugangsberechtigung. Allerdings konnten besonders mit Blick auf die Transferaufgaben signifikante Unterschiede ermittelt werden. Hier konnten sich jene Studierende, die ihre Hochschulzugangsberechtigung am allgemeinbildenden Gymnasium erwarben, akkurater einschätzen als Studierende mit anderweitig erworbener Hochschulzugangsberechtigung.

Hinsichtlich der Motivation konnte gezeigt werden, dass Studierende mit hoher Motivation weniger akkurate Selbsteinschätzungen aufwiesen als Studierende mit geringerer Motivation.

7.2 Diskussion der zentralen Ergebnisse

Die Fähigkeit zur akkuraten Selbsteinschätzung ist potentiell eine wichtige Voraussetzung für selbstreguliertes Lernen. Sie bildet die Grundlage für eine adäquate Auswahl der zu lernenden Inhalte und der aufzuwendenden Lernzeit. Insbesondere im universitären Kontext wird davon ausgegangen, dass dies eine der zentralen Voraussetzungen für den Studienerfolg ist (Köller, 2014; Konegen-Griener, 2002).

In der bisherigen Forschung ist die Akkuratheit von Selbsteinschätzungen in verschiedenen Disziplinen bereits breit erforscht worden (z. B. Metastudie von Blanch-Hartigan (2011) und Längsschnittuntersuchung von Fitzgerald, White, & Gruppen (2003) im medizinischen Kontext oder Brown & Harris (2013) für den schulischen Kontext). Daneben existieren auch zahlreiche Studienbefunde zu den Effekten, die mit der Initiierung von Selbsteinschätzungen verbunden werden (Andrade et al., 2010; Andrade & Valtcheva, 2009; Panadero & Romero, 2014; Panadero et al., 2012; Sadler & Good, 2006). Weniger untersucht wurde bislang der Zusammenhang zwischen der Akkuratheit der Selbsteinschätzung und der Performanz (s. Kapitel 3).

Im Rahmen dieses Forschungsvorhabens wurde die Akkuratheit der Selbsteinschätzung von Studierenden im wirtschaftswissenschaftlichen Kontext betrachtet und auf mögliche Zusammenhänge mit der Performanz untersucht. Zur Bestimmung der Güte der Selbsteinschätzung wurden die Studierenden zu verschiedenen Mess- und Initiierungszeitpunkten im Laufe des Semesters um eine Selbsteinschätzung hinsichtlich ihrer (erwarteten) Performanz bei der Bearbeitung verschiedener Aufgaben gebeten, mit denen unterschiedliche Wissensarten angesprochen wurden.

Die Befragung erfolgte im Rahmen einer wirtschaftswissenschaftlichen Lehrveranstaltung an der Universität Mannheim. Da es sich hierbei um eine Grundlagenveranstaltung im Bachelorstudiengang handelte („Grundlagen des externen Rechnungswesens“), konnten an beiden Messzeitpunkten, verglichen mit den Stichproben der bisher zu diesem Thema durchgeführten Studien, relativ große Stichproben realisiert werden (N = 389 Studierende zum ersten Messzeitpunkt bzw. N = 163 Studierende zum zweiten Messzeitpunkt).

Mit der ersten Forschungsfrage wird der Zusammenhang zwischen akkurater Selbsteinschätzung und Performanz beleuchtet. Hierzu wurde in einem ersten Schritt die Akkuratheit der Selbsteinschätzung detailliert betrachtet. Es wurde davon ausgegangen, dass die Studierenden mehrheitlich akkurate Selbsteinschätzungen aufweisen (s. Hypothese 1_a). Darüber hinaus sollte

sich die Akkuratheit der Selbsteinschätzung zu einem späteren Initiierungs- und Messzeitpunkt verbessern (s. Hypothese 1_b).

Die Annahme, dass die Studierenden mehrheitlich akkurate Selbsteinschätzungen aufweisen, traf an beiden Messzeitpunkten bedingt zu. Es zeigte sich, dass die Akkuratheit der Selbsteinschätzung von ihrem Initiierungszeitpunkt und dem Anspruchsniveau der Aufgaben abhing. Die Selbsteinschätzungen der Studierenden am ersten Messzeitpunkt waren vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) wenig akkurat. Die Studierenden neigten insbesondere hinsichtlich der Text- und Buchungsaufgaben zur Überschätzung ihres eigenen Wissensstandes. Es hat sich gezeigt, dass die Studierenden nur selten in der Lage waren, die Textaufgaben ansatzweise zu beantworten, obwohl sie vor ihrer Bearbeitung geglaubt haben, wichtige Prinzipien und Begriffe wiedergeben zu können. Eventuell messen sie dem weniger Bedeutung zu, da in der Veranstaltung stärker auf die Anwendung der grundlegenden Prinzipien der Buchführung abgezielt wird. Dies könnte eine Erklärung für die geringe Performanz an dieser Stelle darstellen, jedoch nicht erklären, weshalb Studierende dennoch davon ausgingen, die Textaufgaben, die überwiegend Faktenwissen zur Beantwortung erforderten, bearbeiten zu können.

Dagegen war der Anteil der Studierenden, die ihr Abschneiden bzgl. der MC-Aufgaben unterschätzten, relativ hoch. Verhältnismäßig akkurat fiel die Selbsteinschätzung für die Bearbeitung der Transferaufgabe aus. Nach Bearbeitung der Aufgabenstellungen (Initiierungszeitpunkt II) stieg der Anteil der Studierenden mit akkurater Selbsteinschätzung am ersten Messzeitpunkt erwartungsgemäß. Allerdings blieb die Tendenz bestehen, das Abschneiden bzgl. der MC-Aufgaben zu unterschätzen und die Performanz bzgl. der Transferaufgabe akkurater einzuschätzen. Der beobachtete Anstieg in der Akkuratheit der Selbsteinschätzung deckt sich mit den bisherigen Befunden von Maki et al. (1994), die herausfanden, dass Studierende ihre Performanz vor der Aufgabenbearbeitung wenig akkurat prognostizieren konnten. Die Akkuratheit der Selbsteinschätzung verbesserte sich jedoch nach der Aufgabenbearbeitung.

Die Annahme, dass die Selbsteinschätzungen der Studierenden mehrheitlich akkurat sind (Hypothese 1_a), also maximal einen Skalenpunkt von der gezeigten Performanz abweichen, kann zumindest für den zweiten Initiierungszeitpunkt der Selbsteinschätzung für den ersten Messzeitpunkt in Bezug auf das Lösen von Buchungs- und Transferaufgaben bestätigt werden.

Das Ergebnis, dass die Studierenden bei der kognitiv anspruchsvollsten Aufgabenstellung besonders akkurate Selbsteinschätzungen aufweisen, erklärt sich vor dem Hintergrund der tatsächlichen Lösungsquote der Aufgabe. Lediglich etwa jeder zehnte Studierende konnte zum ersten Messzeitpunkt einen adäquaten Lösungsansatz aufzeigen und entsprechend Teilpunkte

erreichen. Insgesamt konnten die Studierenden durchschnittlich nur .28 Punkte von maximal vier möglichen Punkten erzielen. Die Vermutung liegt nahe, dass die Studierenden keinerlei Lösungsansätze erkannten und entsprechend auch die Selbsteinschätzung vornahmen. In diesem Falle haben sie ihr fehlendes Können bei der Bearbeitung der Transferaufgabe erkannt und die Performanz akkurat eingeschätzt. In der Literatur wird kontrovers diskutiert, inwiefern Lernende mit schwächerer Performanz auch über schwächere metakognitive Fähigkeiten verfügen (Kruger & Dunning, 1999). Das in diesem Forschungsvorhaben gezeigte Ergebnis stützt die Ergebnisse von Miller & Geraci (2011), dass Lernende mit schwacher Performanz, sich auch ihrer metakognitiven Defizite bewusst sind. Für diese Annahme spricht auch, dass die Akkuratheit der Selbsteinschätzung zur Lösung von Transferaufgaben bereits zum ersten Initiierungszeitpunkt der Selbsteinschätzung, also vor der Bearbeitung der Transferaufgabe, hoch ausgeprägt war. Die Studierenden konnten also in der Mehrzahl der Fälle antizipieren, dass sie im Umgang mit Transferaufgaben Probleme haben werden. Dies könnte auch mit dem relativ frühen Zeitpunkt im Semester zusammenhängen, zu dem die erste Befragung durchgeführt wurde (sechste Vorlesungswoche). Studierende müssen im externen Rechnungswesen zunächst die Grundlagen der Buchführung verstehen, bevor das erworbene Wissen auf neuartige Aufgabenstellungen angewandt werden kann.

Zum zweiten Messzeitpunkt am Ende des Semesters sollte das Wissen und Können gestiegen sein. Diese Annahme kann mit Blick auf die Performanzwerte zum zweiten Messzeitpunkt bestätigt werden. Fraglich ist, wie sich das gestiegene Wissen und Können der Studierenden auf ihre Selbsteinschätzung ausgewirkt haben könnte. Es konnte eruiert werden, dass sich die Akkuratheit der Selbsteinschätzung zum zweiten Messzeitpunkt tendenziell verschlechterte. Dieser wenig erwartungskonforme Befund (vgl. Hypothese 1_b) liegt vor allem für die Selbsteinschätzung zum zweiten Initiierungszeitpunkt, also nach der Bearbeitung der Aufgaben, vor. Dagegen hat sich die Akkuratheit der Überzeugungen, die Aufgaben korrekt gelöst zu haben, zum zweiten Messzeitpunkt nicht verschlechtert. Dies könnte einerseits darauf zurückzuführen sein, dass die Selbsteinschätzungen hier differenzierter betrachtet werden können, da auf jede Textaufgabe eine Abfrage der Überzeugung folgte. Weiterhin könnte eine weitere Ursache für die diskrepanten Ergebnisse zur Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung im Inhalt der eingesetzten Aufgaben liegen. Diese wurden für den zweiten Messzeitpunkt unverändert vom ersten Messzeitpunkt übernommen, um die Vergleichbarkeit der Angaben zwischen den beiden Messzeitpunkten zu gewährleisten. Die Aufgaben wurden dabei so ausgewählt, dass die Studierenden sie bereits zum ersten Messzeitpunkt mit den in Vorlesung, Übung und Tutorium behandelten Inhalten hätten bewältigen können müssen. Diese Aufgaben

sind jedoch weniger komplex als jene, die Studierende gegen Semesterende zu bewältigen haben. Eventuell bildeten die komplexeren Aufgabenstellungen zum Semesterende die Grundlage für die Selbsteinschätzung der Studierenden vor der Aufgabenbearbeitung. Dass die Aufgaben zum Semesterende eventuell zu leicht gewesen sein könnten, zeigen auch die Tendenzen zu Deckeneffekten, insbesondere bei den MC- und Buchungsaufgaben. Die Performanz könnte aufgrund der verhältnismäßigen Einfachheit der Aufgabenstellungen höher als erwartet ausgefallen sein und sich entsprechend in einem höheren Ausmaß an Unterschätzung niederschlagen, wie für die Lösung der MC-Aufgaben zum zweiten Messzeitpunkt tatsächlich beobachtet werden konnte. Darüber hinaus könnte auch die Aufgabenstellung Unsicherheit und schließlich die Unterschätzung ausgelöst haben, da mehrere Antwortmöglichkeiten richtig sein konnten.

Das Ausmaß der Überschätzung zum zweiten Messzeitpunkt bei der Lösung der Transferaufgabe ist jedoch mit der Annahme, dass die Studierenden zum Ende des Semesters komplexere Aufgabenstellungen als Maßstab für die Selbsteinschätzung herangezogen haben als sie im Rahmen der Befragung lösen mussten, nicht vereinbar. Die Aussagekraft der Ergebnisse ist an dieser Stelle eventuell nicht vollumfänglich gegeben, da die Befragung nur eine Transferaufgabe beinhaltete. Möglicherweise kann die Akkuratheit der Selbsteinschätzung zum zweiten Initiierungszeitpunkt verbessert werden, wenn mehrere Transferaufgaben gestellt werden.

Über beide Messzeitpunkte und alle Initiierungszeitpunkte der Selbsteinschätzung hinweg kann resümiert werden, dass die Studierenden am ersten Messzeitpunkt zu mehrheitlich akkuraten Selbsteinschätzungen neigten. Die Akkuratheit der Selbsteinschätzung stieg dabei zum zweiten Initiierungszeitpunkt. Die Ergebnisse stützen damit die Annahme aus Hypothese 1_b, dass die Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung höher ist als vor der Aufgabenbearbeitung. Die Ergebnisse zum zweiten Messzeitpunkt gestalten sich ambivalenter. Sie deuten zum ersten Initiierungszeitpunkt auf eine insgesamt geringere Akkuratheit der Selbsteinschätzung hin und sind auch zum zweiten Initiierungszeitpunkt nur wenig akkurater. Dagegen liefern die angegebenen Überzeugungen zum Abschneiden der jeweiligen Aufgabenstellungen Anhaltspunkte für eine akkuratere Selbsteinschätzung. Nichtsdestotrotz sind die Ergebnisse besonders in Bezug auf Hypothese 1_b wenig erwartungskonform.

Einen Schwerpunkt des Forschungsvorhabens bildete die Untersuchung des Zusammenhanges zwischen akkurater Selbsteinschätzung und der Performanz. Es wurde davon ausgegangen, dass Studierende mit akkuraterer Selbsteinschätzung zum ersten Messzeitpunkt am zweiten Messzeitpunkt höhere Werte bei der Performanz erzielen (Hypothese 2). Zur Prüfung der zwei-

ten Hypothese wurden die Antworten der Studierenden herangezogen, die an beiden Messzeitpunkten an der Befragung teilgenommen haben ($N = 56$ Studierende). Für diese Teilstichprobe wurde erneut die Akkuratheit der Selbsteinschätzung zum ersten Messzeitpunkt ermittelt. Es zeigte sich, dass diese kaum von jener der Gesamtstichprobe abwich. Die Güte der Selbsteinschätzung war auch hier zum zweiten Initiierungszeitpunkt höher. Im Mittel waren jedoch für die verschiedenen Aufgabentypen etwa 50 % der Selbsteinschätzungen inakkurat.

Entgegen der Erwartungen zeigte sich, dass die Studierenden mit weniger akkuraten Selbsteinschätzungen zum ersten Messzeitpunkt höhere Performanzwerte am zweiten Messzeitpunkt erreichten als Studierende mit akkuraterer Selbsteinschätzung. Dieser Zusammenhang ließ sich für die an verschiedenen Initiierungszeitpunkten gesetzten Selbsteinschätzungen beobachten. Die genauere Analyse des Zusammenhangs ergab, dass die Studierenden, die sich zum ersten Messzeitpunkt unterschätzten, tendenziell höhere Performanzwerte aufweisen als Studierende mit akkurater Selbsteinschätzung oder mit einer Tendenz zur Überschätzung. Die Studierenden, die sich am ersten Messzeitpunkt unterschätzten, erreichten bereits zum ersten Messzeitpunkt höhere Performanzwerte als ihre Kommilitonen mit akkurater Selbsteinschätzung oder Überschätzung. Zum zweiten Messzeitpunkt glichen sich die Performanzwerte jedoch an. Das Ergebnis ist auf den ersten Blick konträr zu bisherigen Forschungsergebnissen (z. B. Thiede et al., 2003; van Loon et al., 2013). Ein Erklärungsansatz für den nicht erwarteten Zusammenhang von inakkurater Selbsteinschätzung und Performanz könnte sein, dass Studierende, die sich eher unterschätzen, mehr Zeit aufbringen und mehr Lernstrategien einsetzen, um eine bessere Performanz zu erreichen. Somit könnte ein Effekt des „Überlernens“ eintreten (Dunlosky & Rawson, 2012). Aus diesem Grund wurden die Items zur Erfassung der eingesetzten Lernstrategien und die Angaben zur aufgewandten Zeit genauer analysiert.

Es wurde erwartet, dass Studierende, die zu Unterschätzung neigen, mehr Zeit und mehr Lernstrategien einsetzen als Studierende mit akkurater Selbsteinschätzung (Hypothese 3 und 4). Insgesamt betrachtet gaben Studierende mit inakkurateren Selbsteinschätzungen signifikant häufiger an, ausgewählte Lernstrategien zu realisieren (z. B. Item LS10: *Ich versuche, die bereitgestellten Übungsaufgaben eigenständig zu lösen.*) als Studierende mit akkuraterer Selbsteinschätzung. Eine weitere Unterscheidung der Fehleinschätzung nach Über- und Unterschätzung ergab, dass Studierende, die am ersten Messzeitpunkt zur Unterschätzung neigten, häufiger Lernstrategien einsetzen als Studierende mit akkurater Selbsteinschätzung. Weniger groß sind die Unterschiede bei Studierende mit der Tendenz, sich zu überschätzen und Studierenden mit akkurater Selbsteinschätzung. Auch die von den Studierenden angegebenen Zeitaufwendungen gehen in die gleiche Richtung. So gaben Studierende, die sich eher unterschätzen an, in einer

durchschnittlichen Woche etwa 300 Minuten für die Bearbeitung der Inhalte in Vorlesung, Übung und Tutorium aufzuwenden, während die übrigen Studierenden etwa ein Drittel weniger Zeit aufbrachten. Der vermutete Effekt des Überlernens könnte diesen Werten zufolge durchaus eine Erklärung auf die wenig erwartungskonformen Ergebnisse zum Zusammenhang von akkurater Selbsteinschätzung und Performanz darstellen.

Ursprüngliches Ziel der Untersuchung war es ferner, zu analysieren, ob es ein quantifizierbares Ausmaß an Über- oder Unterschätzung gibt, welches für die Performanz effektneutral ist (Hypothese 5; Abbildung 28).

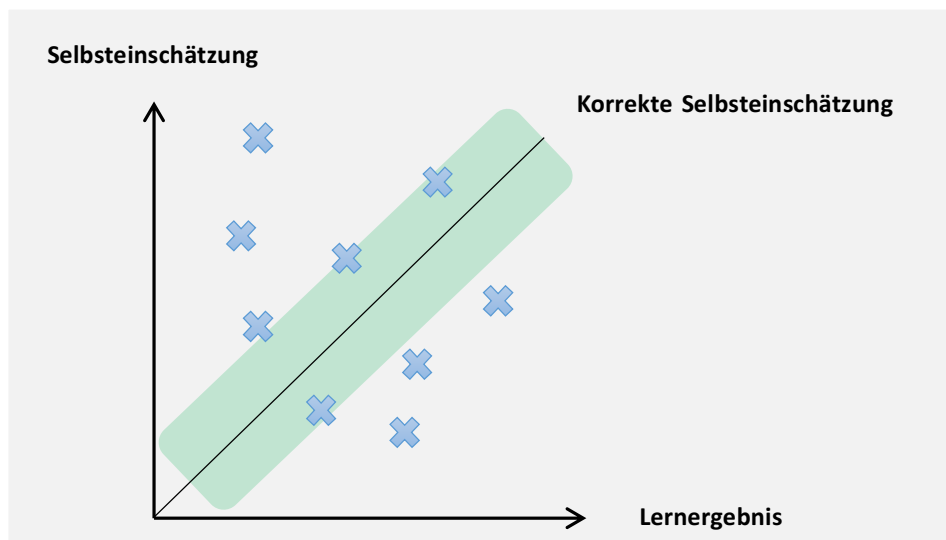


Abbildung 28: Erwarteter Zusammenhang zwischen Fehleinschätzungen und Lernergebnis

In dieser Untersuchung hat sich gezeigt, dass eine Fehleinschätzung nicht nur effektneutral, sondern auch im Zusammenhang mit einer günstigeren Entwicklung der Performanz stehen kann. Aufgrund der hohen Stichprobenmortalität vom ersten zum zweiten Messzeitpunkt gestaltete sich die weitere Analyse dieser Hypothese wenig ergiebig. Die betrachteten Kreuztabellen, in denen die Abweichungen der Selbsteinschätzungen den Performanzwerten gegenübergestellt wurden, zeigen ebenfalls keine klaren Muster, die darauf schließen lassen, dass ein gewisses Ausmaß an Fehleinschätzung die Performanz begünstigt.

Im Rahmen dieses Forschungsprojektes wurden verschiedene Items zur Selbsteinschätzung entwickelt, die sich im Beantwortungsformat (Zuordnung des Wissens zu einer Taxonomiestufe vs. Einschätzung auf Likert-Skala), in der Platzierung innerhalb der Befragung (Selbsteinschätzung vor vs. Selbsteinschätzung nach der Aufgabenbearbeitung) sowie in der inhaltlichen Ausformulierung (Kriterien zur Selbsteinschätzung enthalten vs. keine Kriterien zur Selbsteinschätzung enthalten) unterscheiden. Ziel war es, diejenige Form der Selbsteinschätzung zu eruieren,

die mit besonders akkuraten Selbsteinschätzungen einherging. Dabei wurde davon ausgegangen, dass Studierende sich akkurater einschätzen können, wenn die Selbsteinschätzung an konkreten Kriterien ausgerichtet wird (vgl. Hypothese 6).

Entgegen den Erwartungen stellte sich heraus, dass die Selbsteinschätzungen, die unmittelbar auf eine Aufgabenstellung folgten und die Überzeugung zur Richtigkeit der gegebenen Antwort erfassten, häufig akkurater gewesen sind als die Selbsteinschätzungen vor und nach der Aufgabenbearbeitung (Initiierungszeitpunkt I und II), die nicht unmittelbar auf eine Aufgabenstellung bezogen waren, jedoch Hinweise zu den geforderten Wissensarten enthielten. Offenbar fiel es den Studierenden leichter, ihren Wissensstand akkurat einzuschätzen, wenn die Selbsteinschätzung auf eine konkrete Aufgabe ausgerichtet war. Eine Erklärung hierfür könnte darin bestehen, dass die Anforderungen der hier eingesetzten Aufgaben von hoher Klarheit gewesen sind (z.B. Wiedergabe von Definitionen). Somit war es für die Akkuratheit der Selbsteinschätzung in dieser Untersuchung weniger wichtig, konkrete Kriterien im Rahmen der Items zur Selbsteinschätzung zu formulieren, die Hinweise auf die geforderten Wissensarten gaben. Des Weiteren konnte gezeigt werden, dass die Selbsteinschätzungen über Likert-Skalen leicht akkurater gewesen sind als jene, die eine Zuordnung zu einer Taxonomiestufe vorsahen.

Ausgangspunkt der Überlegungen war die Absicht, ein Instrument zur Selbsteinschätzung zu entwickeln, mit dem akkurate Selbsteinschätzungen gefördert werden können und das aus testökonomischen Gesichtspunkten nicht nur für wenige spezifische Aufgabenstellung in der Veranstaltung „Grundlagen des externen Rechnungswesens“ anwendbar ist, sondern für weite Teile der Veranstaltung. Aufgrund der in Teilen geringen Akkuratheit der Selbsteinschätzungen liegt die Vermutung nahe, dass die Items zur Selbsteinschätzung an Initiierungszeitpunkt I und II eventuell zu allgemein gehalten waren, sodass die Studierenden die dahinterliegenden Inhalte und Aufgaben in Teilen nicht zuordnen konnten.

7.3 Limitationen der Untersuchung und Implikationen für die Forschung

Die Untersuchung liefert Erkenntnisse über das Ausmaß, mit welchem sich Studierende verschiedener wirtschaftswissenschaftlich orientierter Studiengänge akkurat hinsichtlich ihres Wissens- und Könnensstandes im externen Rechnungswesen einschätzen können. Dennoch unterliegt die Ausarbeitung einigen Limitationen, die dazu beigetragen haben können, dass die Ergebnisse in Teilen nicht erwartungskonform gewesen sind.

Hohe Stichprobenmortalität

Wie in Kapitel 3.3.1 dargelegt, wurde bei der Betrachtung des Forschungsstandes festgestellt, dass die Stichprobengröße in vielen Studien eher gering ist. Aus diesem Grund sollten in der vorliegenden Untersuchung die Akkuratheit der Selbsteinschätzung und die Zusammenhänge mit der Performanz für eine größere Kohorte analysiert werden. Deshalb erfolgte die Datenerhebung zu beiden Messzeitpunkten im Rahmen der Übung zur Veranstaltung „Grundlagen des externen Rechnungswesens“, bei der davon ausgegangen wurde, dass sie von Studierenden rege besucht wird. Dieses Vorhaben konnte zum ersten Messzeitpunkt mit einer Stichprobengröße von $N = 389$ Studierende realisiert werden. Zum zweiten Messzeitpunkt war der Rücklauf jedoch deutlich geringer ($N = 163$ Studierende), was sich mit der allgemein sinkenden Teilnahme der Studierenden an der Veranstaltung erklärt. Nichtsdestotrotz konnten mit diesen relativ hohen Fallzahlen aussagekräftige Ergebnisse zur Akkuratheit der Selbsteinschätzung gewonnen werden.

Über die Erfassung der Codes konnten $N = 56$ Studierende identifiziert werden, die an beiden Messzeitpunkt an der Befragung teilgenommen haben. Offenbar haben einige Studierende, die am zweiten Messzeitpunkt an der Befragung teilnahmen, am ersten Messzeitpunkt noch nicht teilgenommen. Weiterhin ist es denkbar, dass die Angabe des Codes zu beiden Messzeitpunkten nicht einheitlich erfolgte und somit Studierende trotz Teilnahme an beiden Messzeitpunkten nicht entsprechend erfasst werden konnten. Die geringe Fallzahl führte dazu, dass die Auswertungen in Teilen nur auf deskriptiver Ebene vorgenommen werden konnten. Dies galt insbesondere bei der Prüfung der Hypothesen 2 bis 4.

Erfassung von Lernstrategien und Lernzeit über Selbsteinschätzungen

Es wurde angenommen, dass die Akkuratheit der Selbsteinschätzung mit der aufgewandten Lernzeit und den realisierten Lernstrategien zusammenhängt. Es konnte gezeigt werden, dass Studierende, die zum ersten Messzeitpunkt zur Unterschätzung neigten, am zweiten Messzeitpunkt angeben, mehr Lernstrategien und Lernzeit aufzuwenden als Studierende mit akkuraterer Selbsteinschätzung. Somit scheint ein Effekt des Überlernens beobachtbar zu sein (Dunlosky & Rawson, 2012). An dieser Stelle ist jedoch diskussionswürdig, inwiefern die auch auf Selbsteinschätzungen basierenden Angaben zu den eingesetzten Lernstrategien sowie der aufgewandten Lernzeit akkurat sind. Möglicherweise kommt es auch hier zu Fehleinschätzungen.

Insbesondere die Erfassung der von Studierenden im Rahmen des Studiums aufgewandten Lernzeit gewann in der Forschung in der Vergangenheit vor allem infolge der Bologna-Reformen, mit denen u. a. eine Umstrukturierung der Studiengänge hin zur Einführung von Bachelor- und Masterstudiengängen einherging, an Bedeutung. Oftmals werden dabei die Zeitaufwendungen in einer typischen Semesterwoche retrospektiv erhoben (z. B. Middendorff, Isserstedt, & Kandulla, 2011; Mishra, 2016). Schulmeister & Metzger (2011) zweifeln die Aussagekraft der mit dieser Methode gewonnenen Ergebnisse an. So könne es zu verzerrten Ergebnissen aufgrund möglicher „Verfälschungseffekte durch Erinnerung“ (Schulmeister & Metzger, 2011, S. 7) oder sozial erwünschter Antworten kommen. Aus diesem Grund stützen sie sich in ihren Untersuchungen auf onlinebasierte Datenerhebungen, die eine tägliche Erfassung der studentischen Tätigkeiten vorsehen. Tatsächlich gelangen sie zur Erkenntnis, dass die dergestalt erfassten Zeitaufwendungen deutlich niedriger sind als in den üblichen retrospektiven Erhebungen.

Im Rahmen dieser Untersuchung ist der Gefahr der verzerrten Ergebnisse hinsichtlich der aufgewandten Lernzeit entgegengewirkt worden, indem die Studierenden verschiedene Items zur aufgewandten Lernzeit zu verschiedenen Zeitpunkten während des Semesters beantworten mussten. Dennoch bleibt der retrospektive Charakter der Erfassung der Lernzeit auch in dieser Untersuchung bestehen, sodass durchaus denkbar ist, dass die ermittelten Lernzeiten höher sind als sie tatsächlich gewesen sind. Fraglich ist jedoch, ob das Phänomen der Überschätzung der aufgewandten Lernzeit gleichermaßen für Studierende gilt, die ihre Performanz unterschätzten, überschätzten oder akkurat einschätzten. In diesem Fall wären die ermittelten absoluten Zeitwerte zwar zu hoch, die Prüfung der Abhängigkeit der aufgewandten Lernzeit von der Akkuratheit der Selbsteinschätzung dennoch möglich. In weiteren Studien könnte der Frage nachgegangen werden, inwiefern eine Fehleinschätzung in der Performanz auch eine Fehleinschätzung der Lernzeit induziert. Hierfür scheint eine Vorgehensweise nach Schulmeister & Metzger (2011) geeignet zu sein, um eine möglichst akkurate Erfassung der Lernzeit zu gewährleisten. Außerdem könnten so auch die Lernstrategien noch detaillierter erfasst werden.

Entwicklung von Items zur Selbsteinschätzung zum ersten und zweiten Initiierungszeitpunkt

Die Items zur Selbsteinschätzung zum ersten und zweiten Initiierungszeitpunkt wurden entwickelt, indem sie sich an den Wissensarten Faktenwissen, konzeptuelles Wissen und prozedurales Wissen orientieren. Zusätzlich wurde ein Item eingefügt, mit dem auf den Wissenstransfer abgezielt wurde. Anschließend wurde untersucht, welche Wissensarten für die Lösung der eingesetzten Aufgaben benötigt werden. Entsprechend wurden die Items zur Selbsteinschätzung den Aufgaben zugeordnet. Durch den Vermerk von kurzen Ankerbeispielen in jedem Item

sollte das Verständnis der Items zur Selbsteinschätzung sichergestellt werden. Allerdings ist nicht auszuschließen, dass die Studierenden Verständnisschwierigkeiten bei den Items zur Selbsteinschätzung zum ersten und zweiten Initiierungszeitpunkt hatten, indem sie nicht erkannten, welche Aufgabentypen den jeweiligen Selbsteinschätzungen zugeordnet sind.

Die Veränderungen in der Selbsteinschätzung zwischen dem ersten und zweiten Initiierungszeitpunkt lassen jedoch den Schluss zu, dass die Studierenden durchaus zu dieser Zuordnung in der Lage gewesen sind. Schließlich sanken die Selbsteinschätzungen zum zweiten Initiierungszeitpunkt, nachdem die Studierenden Schwierigkeiten bei der Aufgabenbearbeitung hatten, wie mit den ermittelten Performanzwerten zum ersten Messzeitpunkt gezeigt werden konnte. Allerdings konnte auch beobachtet werden, dass die ermittelten Überzeugungen, die Aufgabe korrekt gelöst zu haben, unmittelbar nach der Aufgabenbearbeitung in Teilen akkurater waren als die Selbsteinschätzungen an den Initiierungszeitpunkten I und II. Dies könnte einerseits darauf zurückzuführen sein, dass die Selbsteinschätzungen hier differenzierter betrachtet werden können, da bspw. auf jede Textaufgabe eine Abfrage der Überzeugung folgte. Andererseits könnte dieser Befund ein Indiz dafür sein, dass die Studierenden die beabsichtigte Passung von Items zur Selbsteinschätzung und den Aufgabenstellungen nicht immer realisierten. Denkbar ist jedoch auch, dass die auf die spezifischen Aufgaben bezogenen Überzeugungen von den Selbsteinschätzungen, die zwar auf ebendiese Aufgabentypen abzielen, sich jedoch nicht explizit auf sie beziehen, abweichen. In zukünftigen Untersuchungen sollte zunächst das Verständnis der Items zur Selbsteinschätzung gesichert werden (z. B. durch Think-Aloud-Protokolle), um anschließend Aussagen zur Akkuratheit der Selbsteinschätzung zu generieren.

In diesem Kontext ist es ferner denkbar, dass die Passung des Items zur Selbsteinschätzung hinsichtlich des konzeptuellen Wissens („Ich bin in der Lage, die Zusammenhänge zwischen grundlegenden Begriffen und Prinzipien der Buchführung zu erläutern (z. B. Bestimmung erfolgswirksamer Geschäftsvorfälle).“) und der zugeordneten Aufgabenstellung (MC-Aufgaben) nur eingeschränkt gegeben ist. Schließlich wurden mit den MC-Aufgaben keine Erläuterungen gefordert. Eventuell resultiert auch hieraus das hohe Maß an Unterschätzung, das für beide Messzeitpunkte beobachtet werden konnte. Der Hinweis, dass mit dem Item bspw. die Bestimmung erfolgswirksamer Geschäftsvorfälle gemeint gewesen ist, unterstützt allerdings die Annahme, dass Studierende Aufgabe und Selbsteinschätzung einander doch zuordnen konnten. Weiterhin zeigen die angegebenen Überzeugungen unmittelbar nach Bearbeitung der MC-Aufgaben keinen gegenläufigen Trend in der Akkuratheit der Selbsteinschätzung hinsichtlich der MC-Aufgaben.

Kausale Rückschlüsse von Akkuratheit der Selbsteinschätzung auf Performanz nicht möglich

Das dieser Forschungsarbeit zugrundeliegende Untersuchungsdesign lässt nur Aussagen über ungerichtete Zusammenhänge zwischen der Akkuratheit der Selbsteinschätzung und der Performanz zu. Welchen Einfluss bspw. die von den Studierenden eingesetzten Lernstrategien auf diesen Zusammenhang haben, lässt sich nicht abschließend klären. Indizien sprechen dafür, dass mit einer Unterschätzung ein Effekt des Überlernens eintreten kann und so positiv auf das Lernergebnis wirken kann. Um diesen genauer eruieren zu können, sollten die tatsächlichen Fähigkeiten zur adäquaten Aufgabenwahl (*task selection skills*) im Zusammenhang mit der Akkuratheit der Selbsteinschätzung betrachtet werden. Für den schulischen Kontext bestehen hierzu bereits Untersuchungen, die zum Ergebnis kommen, dass die Fähigkeit zur adäquaten Aufgabenwahl trainierbar ist (Kostons et al., 2012) und positive Effekte der adäquaten Aufgabenwahl auf selbstreguliertes Lernen und Performanz konstatieren (Corbalan, Kester, & van Merriënboer, 2008; Kostons et al., 2010). Kostons et al. (2012) entwickelten hierfür Aufgabenraster, wonach sich die von den Lernenden auszuwählenden Aufgaben in Komplexität und Unterstützungsgrad unterscheiden. Eine adäquate Aufgabenwahl läge dann vor, wenn die Lernenden eine Aufgabe aus der nächsthöheren Niveaustufe auswählen, wenn sie die Aufgaben der darunter befindlichen Niveaustufen beherrschen. Dieses Vorgehen könnte auf den universitären und hier auch konkreter auf den wirtschaftswissenschaftlichen Kontext übertragen werden. Im Rahmen dieser Untersuchung war es jedoch vorrangiges Ziel, Zusammenhänge zwischen Akkuratheit der Selbsteinschätzung und der Performanz für eine große Stichprobe zu eruieren.

Literaturverzeichnis

- Adams, T. M. & Ewen, G. W. (2009). The importance of confidence in improving educational outcomes. *25th Annual Conference on Distance Teaching & Learning*, 1–5.
- Alaoutinen, S. (2012). Evaluating the effect of learning style and student background on self-assessment accuracy. *Computer Science Education*, 22(2), 175–198. <http://dx.doi.org/10.1080/08993408.2012.692924>
- Alonso Tapia, J., Panadero, E., & Ruiz, M. (2014). Development and validity of the Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q). *The Spanish Journal of Psychology*, 17, 1–35. <http://doi.org/10.1017/sjp.2014.41>
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57(3), 175–193. <http://dx.doi.org/10.2378/peu2010.art13d>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing - a revision of Bloom's Taxonomy of Educational Objectives* (2nd ed.). Pennsylvania State University, Longman.
- Andrade, H. L., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, 97(1), 21–34. <http://doi.org/10.1080/00220670309596625>
- Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric referenced self assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17(2), 199–214. <http://doi.org/10.1080/09695941003696172>
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting Rubrics to the Test: The Effect of a Model, Criteria Generation, and Rubric-Referenced Self-Assessment on Elementary School Students' Writing. *Educational Measurement: Issues and Practice*, 17(2), 3–13. doi: 10.1111/j.1745-3992.2008.00118.x
- Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *The Journal of Educational Research*, 102(4), 287–302. <http://doi.org/10.3200/JOER.102.4.287-302>
- Andrade, H., & Valtcheva, A. (2009). Promoting Learning and Achievement Through Self-Assessment. *Theory into Practice*, 48(1), 12–19. <http://doi.org/10.1080/00405840802577544>
- Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, 56(1), 45–72. <http://doi.org/10.1007/s11423-007-9067-0>
- Baker, L. (1989). Metacognition, Comprehension Monitoring, and the Adult Reader. *Educational Psychology Review*, 1(1), 3–38.
- Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2), 191–215.
- Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, 61(2), 213–238.

- Barney, S., Khurum, M., Petersen, K., Unterkalmsteiner, M., & Jabangwe, R. (2012). Improving students with rubric-based self-assessment and oral feedback. *IEEE Transactions on Education*, 55(3), 319–325. <http://doi.org/10.1109/TE.2011.2172981>
- Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and Performance of Sales Representatives Test of the Mediating Effects of Goal Setting. *Journal of Applied Psychology*, 78(5), 715–722.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (2000). *Fähigkeit zum selbstregulierten Lernen als fächerübergreifende Kompetenz*. Berlin: MPIB.
- Begg, I. M., Martin, L., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology*, 4(3), 195–218. <http://doi.org/10.1080/09541449208406182>
- Benjamin, A. S. (2007). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. *The Psychology of Learning and Motivation: Skill and Strategy in Memory Use*, 48, 175–223.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Blanch-Hartigan, D. (2011). Medical students' self-assessment of performance: results from three meta-analyses. *Patient Education and Counseling*, 84(1), 3–9. <http://doi.org/10.1016/j.pec.2010.06.037>
- Bloom, B. S. (1984). *Taxonomy of educational objectives*. Boston: Allyn & Bacon.
- Blüthmann, I., Thiel, F., & Wolfgramm, C. (2011). Abbruchtendenzen in den Bachelorstudiengängen - Individuelle Schwierigkeiten oder mangelhafte Studienbedingungen? *Die Hochschule. Journal für Wissenschaft und Bildung*, 1, 110–126.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, 7(2), 161–186. [http://doi.org/10.1016/S0959-4752\(96\)00015-1](http://doi.org/10.1016/S0959-4752(96)00015-1)
- Brand, W., Hofmeister, W., & Tramm, T. (2005). Auf dem Weg zu einem Kompetenzstufenmodell für die berufliche Bildung – Erfahrungen aus dem Projekt ULME. *bwp@ - Berufs- und Wirtschaftspädagogik online*, 8, 1–21.
- Brandstätter, H., Grillich, L., & Farthofer, A. (2006). Prognose des Studienabbruchs. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(3), 121–131. <http://doi.org/10.1026/0049-8637.38.3.121>
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington D.C.: National Academy Press.
- Brown, G. & Harris, L. (2013). Student self-assessment. In J. H. Mc Millian (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 367–393). Los Angeles: SAGE.
- Brown, J. D., & Marshall, M. A. (2001). Self-Esteem and Emotion: Some Thoughts About Feelings. *Personality & Social Psychology Bulletin*, 27(5), 575–584.
- Brown, J. D., & Weiner, B. (1984). Affective consequences of ability versus effort attributions: Controversies, resolutions, and quandaries. *Journal of Educational Psychology*, 76, 146–158. <http://dx.doi.org/10.1037/0022-0663.76.1.146>

- Brown, J. I., Fishco, V. V., & Hanna, G. S. (1993). *Nelson-Denny Reading Test*. Itasca, IL: The Riverside Publishing Company.
- Bruno, J. (1993). Using testing to provide feedback to support instruction: A reexamination of the role of assessment organization. In D. Leclercq & J. Bruno (Eds.), *Item bank: Interactive testing and self-assessment* (pp. 190–209). Berlin: Springer.
- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3), 245–281. <http://doi.org/10.3102/00346543065003245>
- Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances: New results and future directions. *International Journal of Educational Research*, 50(4), 251–256. <http://doi.org/10.1016/j.ijer.2011.08.006>
- Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition & Emotion*, 17(2), 241–261. <http://doi.org/10.1080/02699930244000291>
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer.
- Carver, C. S., & Scheier, M. F. (1990). Origins and Functions of Positive and Negative Affect: A Control-Process View. *Psychological Review*, 97(1), 19–35. doi: 10.1037/0033-295X.97.1.19
- Cassidy, S. (2007). Assessing “inexperienced” students’ ability to self assess: exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education*, 32(3), 313–330. <http://doi.org/10.1080/02602930600896704>
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology*, 33(4), 733–756. <http://doi.org/10.1016/j.cedpsych.2008.02.003>
- Coutts, L., & Rogers, J. (1999). Student Assessment and Standardized Patients - Will the Question never End? *Academic Medicine*, 74(10), 128–130. Retrieved from http://journals.lww.com/academicmedicine/Abstract/1999/10000/Predictors_of_student_self_assessment_accuracy.62.aspx
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350. <http://doi.org/10.1080/03075079912331379935>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation* (5. Auflage). Heidelberg: Springer.
- Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 54, 103–140. doi: 10.1016/B978-0-12-385527-5.00004-8
- Dunlosky, J., & Hertzog, C. (2005). The Self-Monitoring Approach For Effective Learning. *Cognitive Technology*, 10(1), 4–11. Retrieved from <http://psycnet.apa.org/psycinfo/2005-10728-001>
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: SAGE Publications.

- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <http://doi.org/10.1016/j.learninstruc.2011.08.003>
- Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, 59(4), 395–430. <http://doi.org/10.3102/00346543059004395>
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552–564.
- Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education*, 37(7), 645–649. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12834423>
- Flavell, J. H. (1971). First Discussant's Comments: What is Memory Development the Development of? *Human Development*, 14, 272–278.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <http://doi.org/10.1037/0003-066x.34.10.906>
- Flavell, J. H. (1984). Annahmen zum Begriff Metakognition sowie zur Entwicklung von Metakognition. In F. E. Weinert & R. H. Kluwe (Hrsg.), *Metakognition, Motivation und Lernen* (S. 23–31). Stuttgart: Kohlhammer.
- Frye, A. W., Richards, B. F., Bradley, E. W., & Philp, J. R. (1992). Consistency of student's self-assessments in short essay subject matter examinations. *Medical Education*, 26(4), 310–316. doi: 10.1111/j.1365-2923.1992.tb00174.x
- Gernsbacher, M. A., & Varner, K. R. (1988). *The multimedia comprehension battery*. University of Oregon.
- Gillström, A., & Rönnerberg, J. (1995). Comprehension Calibration and Recall Prediction Accuracy of Texts : Reading Skill , Reading Strategies , and Effort. *Journal of Educational Psychology*, 87(4), 545–558.
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 702–718. <http://doi.org/10.1037//0278-7393.11.1-4.702>
- Goforth, D. (1994). Learner Control = Decision Making + Information: A Model and Meta-Analysis. *Journal of Educational Computing Research*, 11(1), 1–26. <http://doi.org/10.2190/X799-7FDG-H0VQ-TH9C>
- Gordon, M. (1991). A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine*, 66(12), 762–769. Retrieved from http://journals.lww.com/academicmedicine/Abstract/1991/12000/A_review_of_the_validity_and_accuracy_of.12.aspx
- Hartwig, M. K., Was, C. A., Isaacson, R. M., & Dunlosky, J. (2012). General knowledge monitoring as a predictor of in-class exam performance. *The British Journal of Educational Psychology*, 82(3), 456–68. <http://doi.org/10.1111/j.2044-8279.2011.02038.x>

- Hasselhorn, M. (1992). Metakognition und Lernen. In G. Nold (Hrsg.), *Lernbedingungen und Lernstrategien: Welche Rolle spielen kognitive Verstehtungsstrukturen?* (S. 35-63). Tübingen: Narr.
- Hasselhorn, M., & Labuhn, A. S. (2008). Metakognition und selbstreguliertes Lernen. In W. Schneider, & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 28–37). Göttingen: Hogrefe.
- Hassmén, P., Sams, M. R., & Hunt, D. P. (1996). Self-Assessment responding and testing methods - effects on performers and observers. *Perceptual and Motor Skills*, 83, 1091–1104.
- Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen*. Hannover: HIS Forum Hochschule.
- Heublein, U., Richter, J., Schmelzer, R., & Sommer, D. (2014). *Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen*. Hannover: HIS Forum Hochschule.
- Hewitt, M. P. (2002). Self-Evaluation Tendencies of Junior High Instrumentalists. *Journal of Research in Music Education*, 50(3), 215–226.
- Hofmeister, W. (2005). Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. *bwp@ - Berufs- und Wirtschaftspädagogik online*, 8, 1–21.
- Hoge, R., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. Retrieved from <http://rer.sagepub.com/content/59/3/297.short>
- Kicken, W., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2008). Scaffolding advice on task selection: a safe path toward self-directed learning in on-demand education. *Journal of Vocational Education & Training*, 60(3), 223–239. <http://doi.org/10.1080/13636820802305561>
- Kim, A., & Leaman, C. (2015). Combat Knowledge Decay in the Workplace. *Talent Development*. Retrieved from: <https://www.td.org/Publications/Magazines/TD/TD-Archive/2015/07/Combat-Knowledge-Decay>
- Kleinknecht, M., Maier, U., Metz, K., & Bohl, T. (2011). Analyse des kognitiven Aufgabenpotentials. *Unterrichtswissenschaft*, 39(4), 328–344.
- Klenowski, V. (1995). Student Self-evaluation Processes in Student-centred Teaching and Learning Contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, 2(2), 145–163. <http://doi.org/10.1080/0969594950020203>
- Köller, O. (2014). Studierfähigkeit und Abitur - empirisch betrachtet. In S. Lin-Klitzing, D. Di Fuccia, & R. Stengl-Jörns (Hrsg.), *Abitur und Studierfähigkeit - Ein interdisziplinärer Dialog* (S. 55–73). Bad Heilbrunn: Klinkhardt.
- Konegen-Griener, C. (2002). Studierfähigkeit und Hochschulzugang. *Forschung und Lehre*, 9, 481–483.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. <http://doi.org/10.1037/0096-3445.126.4.349>

- Koriat, A. (2007). Metacognition and Consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289–325). Cambridge: Cambridge University Press.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for Confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 107–118. <http://doi.org/10.1037/0278-7393.6.2.107>
- Kostons, D., van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Computers & Education*, 54(4), 932–940. <http://doi.org/10.1016/j.compedu.2009.09.025>
- Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121–132. <http://doi.org/10.1016/j.learninstruc.2011.08.004>
- Krathwohl, D. R. (2002). A Revision of Bloom ' s Taxonomy : An Overview. *Theory Into Practice*, 41(4), 212–218.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–34. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10626367>
- Kultusministerkonferenz. (1995). *Weiterentwicklung der Prinzipien der gymnasialen Oberstufe und des Abiturs - Abschlußbericht der von der Kultusministerkonferenz eingesetzten Expertenkommission*. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 464–467. <http://doi.org/10.1037/0278-7393.16.3.464>
- Lichtenstein, S., & Fischhoff, B. (1977). Do Those Who Know More Also Know More about How Much They Know ? *Organizational Behavior and Human Performance*, 20, 159–183.
- Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation*, 23(4), 319–330. [http://doi.org/10.1016/S0191-491X\(97\)86213-X](http://doi.org/10.1016/S0191-491X(97)86213-X)
- Maier, U., Kleinknecht, M., & Metz, K. (2010). Ein fachübergreifendes Kategoriensystem zur Analyse und Konstruktion von Aufgaben. In H. Kiper, W. Meints, S. Peters, S. Schlump, & S. Schmitt (Hrsg.), *Lernaufgaben und Lernmaterialien im kompetenzorientierten Unterricht* (S. 28–43). Stuttgart: Kohlhammer.
- Maki, R. (1998). Metacomprehension of Text: Influence of Absolute Confidence Level on Bias and Accuracy. *Psychology of Learning and Motivation*, 38, 223–248. [https://doi.org/10.1016/S0079-7421\(08\)60188-7](https://doi.org/10.1016/S0079-7421(08)60188-7)
- Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, 1(1), 126–129. <http://doi.org/10.3758/BF03200769>
- Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*, 97(4), 723–731. <http://doi.org/10.1037/0022-0663.97.4.723>

- Marsh, H. W. (1990). The Structure of Academic Self-Concept: The Marsh/Shavelson Model. *Journal of Educational Psychology*, 82(4), 623–636.
- Marsh, H. W., & O'Mara, A. (2008). Reciprocal Effects Between Academic and Attainment Over Seven Adolescent Years: Unidimensional and Multidimensional Perspectives of Self-Concept. *Personality & Social Psychology Bulletin*, 34, 542–552. <http://doi.org/10.1177/0146167207312313>
- Marshall, M., & Brown, J. (2006). Emotional reactions to achievement outcomes: Is it really best to expect the worst? *Cognition & Emotion*, 20(1), 43–63. <http://doi.org/10.1080/02699930500215116>
- McDonald, B., & Boud, D. (2003). The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209–220. <http://doi.org/http://dx.doi.org/10.1080/0969594032000121289>
- Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision Affect Theory Emotional reactions to the outcomes of risky options. *Psychological Science*, 8(6), 423–429.
- Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, 38(4), 441–51. <http://doi.org/10.3758/MC.38.4.441>
- Merrill, M. D. (1980). Learner Control in Computer Based Learning. *Computers and Education*, 4, 77–95.
- Metzger, C., Waibel, R., Henning, C., Hodel, M., & Luzi, R. (1993). Anspruchsniveau von Lernzielen und Prüfungen im kognitiven Bereich. *Studien und Berichte des IWP*, 10, St. Gallen: Institut für Wirtschaftspädagogik.
- Middendorff, E., Isserstedt, W., & Kandulla, M. (2011). *Studierende im Bachelor-Studium 2009: Ergebnisse der 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem*. Bonn: BMBF.
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 37(2), 502–6. <http://doi.org/10.1037/a0021802>
- Mishra, S. (2016). Erwerbstätigkeit und Zeitbudget der Studierenden im Europäischen Hochschulraum. *EUROSTUDENT Kurzdossier*. Hannover: DZHW.
- Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias? *International Journal of Educational Research*, 50(4), 230–240. <http://doi.org/10.1016/j.ijer.2011.08.004>
- Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–173. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079742108600535/pdf?md5=57cd82e8a8dc29c3252c7322471976e1&pid=1-s2.0-S0079742108600535-main.pdf&_valck=1
- Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, 10, 257–260.

- Nelson, T. O., Leonesio, R. J., Landwehr, R. S., & Narens, L. (1986). A Comparison of Three Predictors of an Individual's Memory Performance : The Individual's Feeling of Knowing Versus the Normative Feeling of Knowing Versus Base-Rate Item Difficulty. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12(2), 279–287.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, 1(2), 159–179. <http://doi.org/10.1007/s10409-006-9595-6>
- Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD). (2014). *PISA 2012 Results: Creative Problem Solving* (Vol. V). <http://doi.org/10.1787/9789264208070-en>
- Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development*, 52(3), 5–22. <http://doi.org/10.1007/BF02504672>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133–148. <http://doi.org/10.1080/0969594X.2013.877872>
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806–813. <http://doi.org/10.1016/j.lindif.2012.04.007>
- Pintrich, P. R. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Retrieved from <http://www.ilo.uva.nl/projecten/Gert/Teaching/VietnamCourseDec2006ResearchMethodology/Pintrich.pdf>
- Preckel, D. (2006). Fallstudien – unverzichtbar und lehrreich. *Netzwerk*, 4, 32–33.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (1999). Effects of Self-Evaluation Training on Narrative Writing. *Assessing Writing*, 6(1), 107–132.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student Self-Evaluation in Grade 5 – 6 Mathematics Effects on Problem- Solving Achievement. *Educational Assessment*, 8(1), 43–59.
- Roth, A., Ogrin, S., & Schmitz, B. (2015). Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*. <http://doi.org/10.1007/s11092-015-9229-2>
- Ruddel, R. B., & Speaker, R. B. (1985). The interactive reading process: A model. In H. Singler & R. B. Ruddel (Eds.), *Theoretical Models and Processes of Reading* (3rd ed.) (pp. 751–793). Newark, Del.: International Reading Association.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sadler, P., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1), 1–31. http://doi.org/10.1207/s15326977ea1101_1
- Saunders, T. C., & Holahan, J. M. (1997). Criteria-Specific Rating Scales in the Evaluation of High School Instrumental Performance. *Journal of Research in Music Education*, 45(2), 259–272.

- Schmitz, B. (2001). Self-Monitoring zur Unterstützung des Transfers einer Schulung in Selbstregulation für Studierende. *Zeitschrift für Pädagogische Psychologie*, 15(3/4), 181–197. <http://doi.org/https://doi.org/10.1024/1010-0652.15.34.181>
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26, 113–125.
- Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Schulmeister, R., & Metzger, C. (2011). Zum Projekt ZEITLast oder: eine realistischere Zeitbudgetforschung zum Bachelorstudium. *Wissenswert*, 2, 5–11.
- Shrauger, J. S. (1975). Responses to Evaluation as a Function of Initial Self-Perceptions. *Psychological Bulletin*, 82(4), 581–596.
- Simons, J. (1992). Lernen, selbstständig zu lernen - ein Rahmenmodell. In H. Mandl (Hrsg.), *Lern- und Denkstrategien: Analyse und Intervention* (S. 251–264). Göttingen: Hogrefe.
- Sitzmann, T., & Johnson, S. K. (2012). When is ignorance bliss? The effects of inaccurate self-assessments of knowledge on learning and attrition. *Organizational Behavior and Human Decision Processes*, 117(1), 192–207. <http://doi.org/10.1016/j.obhdp.2011.11.004>
- Snyder, K. E., Nietfeld, J. L., & Linnenbrink-Garcia, L. (2011). Giftedness and Metacognition: A Short-Term Longitudinal Investigation of Metacognitive Monitoring in the Classroom. *Gifted Child Quarterly*, 55(3), 181–193. <http://doi.org/10.1177/0016986211412769>
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <http://doi.org/10.1037/a0027627>
- Swann, W. B., Griffin, J. J., Predmore, S. C., & Gaines, B. (1987). The Cognitive-Affective Crossfire: When Self-Consistency Confronts Self-Enhancement. *Journal of Personality and Social Psychology*, 52(5), 881–889.
- Sweller, J., Merrienboer, J. J. G. Van, & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. doi: 10.1023/a:1022193728205
- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6(4), 662–667.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66–73. <http://doi.org/10.1037/0022-0663.95.1.66>
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 1024–1037. <http://doi.org/10.1037/0278-7393.25.4.1024>
- Trapmann, S., Hell, B., & Weigand, S. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21(1), 11–27. <http://doi.org/10.1024/1010-0652.21.1.11>
- Universität Mannheim. (2017). *Modulkatalog Grundlagen der Betriebswirtschaftslehre und Wirtschaftsinformatik für Studierende anderer Fachrichtungen*. Universität Mannheim.

- van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*, 24, 15–25. <http://doi.org/10.1016/j.learninstruc.2012.08.005>
- Vygotsky, L. S. (1978). *Mind in society: The development of Higher Psychological Processes*. Cambridge: Harvard University Press.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96(3), 465–490.
- Weiner, B. (1986). *An attributional theory of achievement motivation and emotion*. New York: Springer.
- Weinstein, C. E., & Palmer, D. R. (2002). *Learning and study strategies inventory LASSI user's manual* (2nd edition). Clearwater: H&H Publishing Company.
- Weinstein, C. E., Zimmermann, S. A., & Palmer, D. R. (1988). Assessing Learning Strategies: The Design and Development of the Lassi. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and Study Strategies: Issues in Assessment, Instruction, and Evaluation* (pp. 25–40). San Diego: Academic Press.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, New York: Erlbaum.
- Wolter, S. C., Diem, A., & Messer, D. (2013). *Studienabbrüche an Schweizer Universitäten*. Aarau: SKBF.
- Wüstemann, J. (2015). *Buchführung case by case*. Frankfurt am Main: dfv Mediengruppe.
- Zechmeister, E. B., & Bennett, D. J. (1991). How easy is it to judge ease of learning? *Bulletin of the Psychonomic Society*, 29(1), 36–38. <http://doi.org/10.3758/BF03334762>
- Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1–19). New York: Guilford Publications.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). San Diego: Academic Press.
- Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner : An Overview. *Theory Into Practice*, 41(2), 64–70. <http://doi.org/10.1207/s15430421tip4102>
- Zimmerman, B. J., & Campillo, M. (2003). Motivating Self-Regulated Problem Solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The Psychology of Problem Solving* (pp. 233–262). Cambridge: University Press. <http://doi.org/10.1017/CBO9780511615771.009>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-Regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York: Routledge.

Anhang

Anhang A1	Synopse der Studien zu den Effekten (in-)akkuratere Selbsteinschätzungen auf das Lernen	153
Anhang A2	Codebuch	176
Anhang A3	Mittelwertvergleich der Selbsteinschätzungen der Gesamtstichprobe	188
Anhang A4	Mittelwertvergleich der Selbsteinschätzungen vor der Aufgabenbearbeitung für beide Messzeitpunkte ($N = 56$).....	189
Anhang A5	Mittelwertvergleich der Selbsteinschätzungen nach der Aufgabenbearbeitung für beide Messzeitpunkte ($49 \leq n \leq 56$).....	189
Anhang A6	Mittelwertvergleich der Überzeugungen für beide Messzeitpunkte ($27 \leq n \leq 52$).....	190
Anhang A7	Übersicht der erreichten Punktzahl in Abhängigkeit von der Selbsteinschätzung (SE121)	191
Anhang A8	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt	192
Anhang A9	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt.....	192
Anhang A10	Übersicht über Zusammenhänge zwischen Überzeugungen und der Performanz zum ersten Messzeitpunkt.....	193
Anhang A11	Übersicht der erreichten Punktzahl in Abhängigkeit von der Selbsteinschätzung (SE221)	194
Anhang A12	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum zweiten Messzeitpunkt	195
Anhang A13	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum zweiten Messzeitpunkt.....	195
Anhang A14	Übersicht über Zusammenhänge zwischen Überzeugungen und der Performanz zum zweiten Messzeitpunkt.....	196
Anhang A15	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit des Geschlechts	197
Anhang A16	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit des Geschlechts.....	198
Anhang A17	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit der Hochschulzugangsberechtigung.....	199
Anhang A18	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit der Hochschulzugangsberechtigung	200

Anhang A19	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit der Motivation	201
Anhang A20	Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit der Motivation	202
Anhang A21	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt	203
Anhang A22	Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt.....	203
Anhang A23	Mittelwertvergleich der Performanz in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen insgesamt	204
Anhang A24	Deskriptive Ergebnisse zu den Items der Lernstrategien und Lernzeit zu beiden Messzeitpunkten	205
Anhang A25	Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) am ersten Messzeitpunkt.....	206
Anhang A26	Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung (Initiierungszeitpunkt II) am ersten Messzeitpunkt.....	207
Anhang A27	Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) am zweiten Messzeitpunkt.....	208
Anhang A28	Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung (Initiierungszeitpunkt II) am zweiten Messzeitpunkt.....	209

Anhang A1 Synopse der Studien zu den Effekten (in-)akkurater Selbsteinschätzungen auf das Lernen

Anhang A1.1 Übersicht der berücksichtigten Studien

Nr.	Autoren & Titel der Studie
1	Begg, I. M., Martin, L., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? <i>European Journal of Cognitive Psychology</i> , 4(3), 195–218. http://doi.org/10.1080/09541449208406182
2	Coutts, L., & Rogers, J. (1999). Student Assessment and Standardized Patients - Will the Question never End? <i>Academic Medicine</i> , 74(10), 128–130.
3	Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. <i>Learning and Instruction</i> , 22(4), 271–280. http://doi.org/10.1016/j.learninstruc.2011.08.003
4	Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. <i>Journal of Experimental Psychology: Learning, Memory, and Cognition</i> , 11(4), 702–718. http://doi.org/10.1037//0278-7393.11.1-4.702
5	Hartwig, M. K., Was, C., Isaacson, R. M., & Dunlosky, J. (2012). General knowledge monitoring as a predictor of in-class exam performance. <i>The British Journal of Educational Psychology</i> , 82(3), 456–68. http://doi.org/10.1111/j.2044-8279.2011.02038.x
6	Hewitt, M. P. (2002). Self-Evaluation Tendencies of Junior High Instrumentalists. <i>Journal of Research in Music Education</i> , 50(3), 215–226.
7	Kostons, D., van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. <i>Computers & Education</i> , 54(4), 932–940. http://doi.org/10.1016/j.compedu.2009.09.025
8	Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. <i>Learning and Instruction</i> , 22(2), 121–132. http://doi.org/10.1016/j.learninstruc.2011.08.004
9	Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. <i>Psychonomic Bulletin & Review</i> , 1(1), 126–129. http://doi.org/10.3758/BF03200769
10	Marshall, M., & Brown, J. (2006). Emotional reactions to achievement outcomes: Is it really best to expect the worst? <i>Cognition & Emotion</i> , 20(1), 43–63. http://doi.org/10.1080/02699930500215116

Nr.	Autoren & Titel der Studie
11	Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: stability and generality in the learning process and predictive validity for learning outcome. <i>Memory & Cognition</i> , 38(4), 441–451. http://doi.org/10.3758/MC.38.4.441
12	Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias? <i>International Journal of Educational Research</i> , 50(4), 230–240. http://doi.org/10.1016/j.ijer.2011.08.004
13	Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. <i>The Journal of Experimental Education</i> , 74(1), 7–28.
14	Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. <i>Educational Technology Research and Development</i> , 52(3), 5–22. http://doi.org/10.1007/BF02504672
15	Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. <i>Assessment in Education: Principles, Policy & Practice</i> , 21(2), 133–148. http://doi.org/10.1080/0969594X.2013.877872
16	Sitzmann, T., & Johnson, S. K. (2012). When is ignorance bliss? The effects of inaccurate self-assessments of knowledge on learning and attrition. <i>Organizational Behavior and Human Decision Processes</i> , 117(1), 192–207. http://doi.org/10.1016/j.obhdp.2011.11.004
17	Snyder, K. E., Nietfeld, J. L., & Linnenbrink-Garcia, L. (2011). Giftedness and Metacognition: A Short-Term Longitudinal Investigation of Metacognitive Monitoring in the Classroom. <i>Gifted Child Quarterly</i> , 55(3), 181–193. http://doi.org/10.1177/0016986211412769
18	Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. <i>Journal of Educational Psychology</i> , 95(1), 66–73. http://doi.org/10.1037/0022-0663.95.1.66
19	van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. <i>Learning and Instruction</i> , 24, 15–25. http://doi.org/10.1016/j.learninstruc.2012.08.005

Anhang A1.2 Details zu berücksichtigten Studien

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
1	Begg, Martin, & Needham (1992)	<ul style="list-style-type: none"> ➤ Kein expliziter theoretischer Bezugsrahmen skizziert ➤ Wichtigkeit des Self-Monitoring für selbstgesteuerte Kontrollprozesse (z. B. Verteilung der Lernzeit) hervorgehoben (Leonesio & Nelson, 1990) 	Studie 1 <ul style="list-style-type: none"> ➤ Welche Effekte haben Cue-Reviews und Pair-Reviews beim Lernen von Wortpaaren auf das Lernergebnis? ➤ Hat das Monitoring (Selbsteinschätzung) einen Effekt auf das Lernen von nicht gewussten Wortpaaren? 	Studie 1 <ul style="list-style-type: none"> ➤ N = 43 Studierende eines Einführungskurses in Psychologie ➤ 2 TG (cue review & pair review): <ul style="list-style-type: none"> i) cue-review (Railroad → ?) ii) pair-review (Railroad – Mother) ➤ Lernenden werden 96 Wortpaare (z. B. Railroad – Mother) jeweils für vier Sekunden präsentiert ➤ Anschließend Selbsteinschätzung auf fünfstufiger Likert-Skala, ob Wortpaar wiedergegeben werden kann: <ul style="list-style-type: none"> i) cue-review (Railroad → ?) ii) pair-review (Railroad – Mother) 	Studie 1 <ul style="list-style-type: none"> ➤ Lernende mit Pair-Review mit besserem Recall als Cue-Review ➤ Selbsteinschätzung nicht hilfreich für Lernende i. S. einer effizienten Allokation der Lernmöglichkeiten ➤ Akkuratheit der Selbsteinschätzung für Cue-Review besser als für Pair-Review
		<ul style="list-style-type: none"> ➤ Feeling of Knowing (Nelson, Leonesio, Landwehr, & Narens, 1986) 	Studie 2 <ul style="list-style-type: none"> ➤ Welchen Effekt hat ein Pair- und Cue-Review gelerntes Wortpaar i. Vgl. zur einfachen Wiederholung? ➤ Werden wiederholte Items besser erinnert, wenn vor der Wiederholung eine Selbsteinschätzung erfolgt? 	Studie 2 <ul style="list-style-type: none"> ➤ N = 86 Studierende in Einführungskurs der Psychologie ➤ 100 Wortpaare in drei Etappen (Anfangsliste, Reviewliste, finale Liste), abschließend Test 	Studie 2 <ul style="list-style-type: none"> ➤ Monitoring ohne Einfluss auf die Performanz

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
			<ul style="list-style-type: none"> ➤ Werden Items, die Lernende vor der Wiederholung als nicht wissend einstufen, eher erinnert? ➤ Es wird erwartet, dass sämtliche Fragestellungen verneint werden müssen. 		
2	Coutts & Rogers (1999)	<ul style="list-style-type: none"> ➤ Kein theoretischer Bezugsrahmen skizziert, es wird lediglich auf Studien zur Akkuratheit der Selbsteinschätzung im medizinischen Kontext verwiesen (z. B. Frye, Richards, Bradley, & Philp, 1992; Gordon, 1991) 	<ul style="list-style-type: none"> ➤ Welche Prädiktoren für eine akkurate Selbsteinschätzung lassen sich identifizieren? ➤ Worin unterscheiden sich Lernende, die sich über- bzw. unterschätzen? ➤ Welche Zusammenhänge bestehen zwischen Akkuratheit der Selbsteinschätzung und der Performanz? 	<ul style="list-style-type: none"> ➤ N = 420 Studierende der Medizin ➤ Bearbeitung von medizinischen Fallstudien (standardisierte Patienten) ➤ Einsatz verschiedener Skalen (Risikobereitschaft, Entscheidung, Umgang mit Ungewissheit) ➤ Selbsteinschätzung der Performanz für jede Fallstudie auf einer fünfstufigen Skala ➤ Bestimmung der Diskrepanzen zur tatsächlichen Performanz und Entwicklung eines Bias-Index ➤ Bildung dreier Gruppen (Über-/Unterschätzung und akkurate Selbsteinschätzung), cut off bei 20 % Diskrepanz (analog zur fünfstufigen Skala) 	<ul style="list-style-type: none"> ➤ 86.8 % der Lernenden innerhalb des als akkurat erachteten Intervalls +/- 20 %, Unterschätzung bei 7.9 % und Überschätzung bei 5.3 % der Probanden ➤ Die in der Studie berücksichtigten zehn Prädiktoren erklären nur 15 % der Varianz der Akkuratheit der Selbsteinschätzung (signifikant sind Humanismus, Risikobereitschaft, Zahl ähnlicher gesehener Patienten, Geschlecht) ➤ Lernende mit Überschätzung tendieren zu höherer Risikobereitschaft als Unterschätzer ➤ Gruppen der akkuraten Selbsteinschätzer und der Unterschätzer weisen hö-

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					here Werte in der Performanz auf als die Gruppe der Überschätzer
3	Dunlosky & Rawson (2012)	<ul style="list-style-type: none"> ➤ Theorie der Selbstregulation: Output des Lernens beeinflusst Monitoring und Monitoring wiederum die Kontrollprozesse des Lernens (Boekaerts, 1997; Dunlosky & Metcalfe, 2009; Flavell, 1979; Nelson & Narens, 1990) ➤ Accuracy-Influences-Memory-Hypothese (AIM): Akkuratheit im Monitoring führt zu effektiverem Lernen und besseren Lernergebnissen (Benjamin, 2007; Dunlosky & Ariel, 2011) ➔ Annahme: Monitoring wird effektiv zur Kontrolle genutzt (Thiede, 1999) 	<ul style="list-style-type: none"> ➤ Welche Zusammenhänge bestehen zwischen der absoluten Akkuratheit der Selbsteinschätzung und der Performanz in einem abschließenden Test? ➤ Lernende, die sich unterschätzen, können von einem Effekt des Überlernens profitieren; Überschätzung hat aber in gleichem Ausmaß einen negativen Effekt auf das Lernen ➤ Zwei Studien: <ol style="list-style-type: none"> i) Manipulation der Akkuratheit via Gabe verschiedener Prompts ii) Untersuchung individueller Unterschiede in der Akkuratheit der Selbsteinschätzung 	Studie 1 <ul style="list-style-type: none"> ➤ N = 48 Studierende der Psychologie ➤ Prozedere: Lernen von Definitionen, Recall der gelernten Definitionen, Selbsteinschätzung ➤ TG (idea-unit-standard: Zeigen der Antworten mit Hinweisen zur korrekten Antwort, Checkliste, danach Selbsteinschätzung); KG (no standard-group: Zeigen der Antwort, danach Selbsteinschätzung) ➤ Texte zu psychologischen Konzepten mit Definitionen ➤ Selbsteinschätzung auf dreistufiger Skala (no credit, partial credit, full credit) ➤ Computergestützte Auswahl der Definitionen zur Wiedergabe/Wiederholung ➤ Dauer der Wiederholungsphasen von Lernenden gesteuert ➤ Item wird aus der Liste der zu wiederholenden Definitionen genommen, wenn Antwort hierzu drei Mal für korrekt befunden wurde 	Studie 1 <ul style="list-style-type: none"> ➤ Testergebnisse für TG > KG ➤ Ausmaß der Unterschätzung in beiden Gruppen gering ausgeprägt und nicht signifikant verschieden ➤ Überschätzung für KG signifikant stärker ausgeprägt als für TG ➤ Überschätzung führt zur früheren Beendigung der Übungseinheiten und geringerer Performanz ➤ Keine Unterschiede in der Lernzeit zwischen TG und KG; erklärt nicht Unterschiede in der Performanz

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				Studie 2 <ul style="list-style-type: none"> ➤ Nur noch idea-unit-standard-Gruppe ➤ N = 158 Studierende der Psychologie ➤ Ablauf weitgehend identisch mit Studie 1, aber keine Selbsteinschätzung nach idea-unit-Einschätzung mehr nötig 	Studie 2 <ul style="list-style-type: none"> ➤ Je stärker die Überschätzung ausgeprägt ist, desto schlechter ist die Performanz im abschließenden Test ➤ Korrelation zwischen Überschätzung und Performanz bei $r = -.75$ ➤ Korrelation zwischen Überschätzung und Performanz in letzter Übungsrunde bei $r = -.93$ ➤ Korrelation zwischen finaler Performanz in der Übungsphase und jener beim Test bei $r = .71$ ➤ Lernzeit unabhängig von Akkuratheit der Selbsteinschätzung
4	Glenberg & Epstein (1985)	<ul style="list-style-type: none"> ➤ Kein theoretischer Bezugsrahmen skizziert, es wird lediglich auf Studien verwiesen, die Problematik des <i>illusion of knowing</i> aufgreifen (Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein & Fischhoff, 1977) 	<ul style="list-style-type: none"> ➤ Welchen Effekt hat eine Variation des Zeitabstands zwischen Bearbeitung eines Textes und der Selbsteinschätzung auf diese? 	Studie 1 <ul style="list-style-type: none"> ➤ N = 85 Studierende in Einführungskurs Psychologie ➤ 2 TG (immediate & delayed self-assessment) ➤ Bearbeitung von 15 Kurztexten, im Anschluss Selbsteinschätzung (immediate vs. delayed) auf sechsstufiger Skala 	Studie 1 <ul style="list-style-type: none"> ➤ Korrelationen, Bestimmung des Confidence Judgements Accuracy Quotienten und Erstellung der Kalibrierungskurven zeigen, dass Lernende kaum in der Lage sind, zu unterscheiden, welche Aufgaben verstanden wurden und welche nicht

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				<ul style="list-style-type: none"> ➤ 15 Testaufgaben bzgl. des Verständnisses der Texte im Format w/f-Aussagen mit unterschiedlichen kognitiven Anspruchsniveaus ➤ Lernende erhalten keine Rückmeldung während der Testung zur Akkuratheit der Selbsteinschätzung 	
			<ul style="list-style-type: none"> ➤ Wie verhält sich der Zusammenhang zwischen Selbsteinschätzung und Testperformanz, wenn Lernende zunächst mit dem Testverfahren vertraut gemacht werden? 	Studie 2 <ul style="list-style-type: none"> ➤ N = 60 Studierende eines Einführungskurses in die Psychologie ➤ 2 TG (familiarization (Übung zum Prozedere mit drei Beispielen zu Beginn) vs. non-familiarization (bloßes Lesen der Beispiele)) 	Studie 2 <ul style="list-style-type: none"> ➤ Ähnliche Ergebnisse wie in Studie 1; Vertrautheit mit Testformat nicht ausschlaggebend
			<ul style="list-style-type: none"> ➤ Wie entwickelt sich die Kalibrierung, wenn Lernende auf Erfahrung zurückgreifen können? 	Studie 3 <ul style="list-style-type: none"> ➤ N = 39 Studierende ➤ Zyklus von fünf Etappen: <ol style="list-style-type: none"> Einschätzung nach Lernen, Testaufgabe zum gelesenen Text, Confidence Judgements bzgl. des Tests, Überdenken der SE aus i), neue Testaufgabe (Recalibration) ➤ Bestimmung von 3 Kalibrierungsscores: Zusammenhang zwischen Selbsteinschätzung zu Beginn & Performanz im ersten Test, Zusammenhang zwischen Performanz im ersten Test & Selbsteinschätzung nach Test, Zusammenhang zwischen 	Studie 3 <ul style="list-style-type: none"> ➤ Zu Beginn gering ausgeprägte Fähigkeit zur SE ➤ Kalibrierung der Performanz bereits besser ($r = .23$) ➤ Rekalibrierung besser als bei Einschätzung zu Beginn ($r = .19$) ➤ Insgesamt keine Veränderung über die verschiedenen Zeitpunkte hinweg in Selbsteinschätzung → akkuratere SE der Über-/Unterschätzer

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				erneuter Selbsteinschätzung & Performanz im zweiten Test	<ul style="list-style-type: none"> ➤ Confidence-Judgement Accuracy Quotient bestätigt Ergebnisse, jedoch keine signifikante Verbesserung zwischen Kalibrierung zu Beginn und Rekalibrierung ➤ Kalibrierungskurven bestätigen Befunde
5	Hartwig, Was, Isaacson, & Dunlosky (2012)	<ul style="list-style-type: none"> ➤ Theorien selbstregulierten Lernens (Dunlosky & Metcalfe, 2009) ➤ Selbstreguliertes Lernen nach Winne & Hadwin (1998) 	<ul style="list-style-type: none"> ➤ Wird die Performanz der Lernenden von der Fähigkeit zur akkuraten Selbsteinschätzung irrelevanter Inhalte vorhergesagt? ➤ Wird die Performanz stärker von der Akkuratheit des Monitorings hinsichtlich inhaltlich verwandter, jedoch nicht relevanter Materialien für die finale Testung vorhergesagt als durch unverwandte Materialien? 	<ul style="list-style-type: none"> ➤ N = 379 Studierende im Einführungskurs „Pädagogische Psychologie“ ➤ Akkuratheit des Monitorings vorab gemessen via binärer Selbsteinschätzung (ja/nein) und Verständnistest (MC mit fünf Alternativen) zu 50 Vokabeln (davon 33 inhaltlich verwandt mit finaler Testung) ➤ Akkuratheit bestimmt aus <i>hit</i> (+/+), <i>false alarm</i> (+/-), <i>miss</i> (-/+) und <i>correct rejection</i> (-/-); anschließend Bestimmung von γ-Korrelationen für jeden Probanden ➤ Abschließender Test bestehend aus 100 Items zu Inhalten des Semesters in verschiedenen Niveaustufen ➤ Vokabeltest als Indikator für verbale Fähigkeit herangezogen und fortan kontrolliert 	<ul style="list-style-type: none"> ➤ Korrelation zwischen Akkuratheit des Monitorings und Ergebnis beim Abschlusstest bei $r = .26$ ➤ Akkuratheit des Monitorings erklärt 6.8 % der Varianz der Performanz ➤ Lernende mit höheren Akkuratheitswerten weisen höhere Performanzwerte auf ➤ Zusammenhang gilt auch für gänzlich unverwandte Items ($r = .17$) ➤ Zusammenhänge auch für andere Prüfungen nachgewiesen ➤ Positiver Zusammenhang zwischen verbaler Fähigkeit und Testergebnis ($r = .39$), bei Kontrolle bleibt der beobachtete Zusammen-

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					menhang für die Akkuratheit des Monitorings jedoch bestehen
6	Hewitt (2002)	<ul style="list-style-type: none"> ➤ Selbstreguliertes Lernen nach Zimmerman (2000): zyklischer Prozess bestehend aus Forethought, control & Self-Reflection-Phase ➤ <i>Self-evaluation</i> wesentlicher Bestandteil der <i>Self-Reflection</i>-Phase (Zimmerman, 1998) 	<ul style="list-style-type: none"> ➤ Wie entwickelt sich die Selbsteinschätzung im Zeitverlauf? ➤ Welchen Effekt hat die Modellierung der Stücke (Vorspiel) auf die Akkuratheit der Selbsteinschätzung? ➤ Welche Zusammenhänge bestehen zwischen der Akkuratheit der Selbsteinschätzung und der Performanz? 	<ul style="list-style-type: none"> ➤ N = 41 Lernende im Fach Musik (7. – 9. Klassen) ➤ Messung der Performanz & Selbsteinschätzung mit <i>Woodwind Brass Solo Evaluation Form</i> von Saunders & Holahan (1997) ➤ sechswöchiger regulärer Unterricht, in welchem Umgang mit Instrument geübt wird ➤ 2 TG (Hören von richtigem Spiel vs. kein Hören) ➤ Am Ende der sechs Wochen folgt Selbsteinschätzung der Performanz in sieben Kategorien 	<ul style="list-style-type: none"> ➤ Selbsteinschätzung steigt im Zeitverlauf unabhängig von der TG ➤ Das Treatment hat keinen Effekt auf die Akkuratheit der Selbsteinschätzung ➤ Akkuratheit der Selbsteinschätzung steigt nicht im Zeitverlauf; Tendenz zur Überschätzung ➤ Keine signifikanten Zusammenhänge zwischen Experteneinschätzung (Performanz) und Selbsteinschätzung der Lernenden im Pretest; Posttest zeigt lediglich signifikante Zusammenhänge für zwei Teilbereiche
7	Kostons, van Gog, & Paas (2010)	<ul style="list-style-type: none"> ➤ Learner-Controlled-Instruction (LCI): Wahl der eigenen Lernaktivitäten auf Basis der eigenen Lernbedürfnisse (Merrill, 1980) ➤ Selbstgesteuertes Lernen aber oftmals nicht effektiv: Ursache evtl. mangelnde Fähigkeit zur Selbsteinschätzung 	<ul style="list-style-type: none"> ➤ Welche Unterschiede bestehen zwischen effektiven und ineffektiven Lernenden hinsichtlich der Selbsteinschätzung und der Aufgabenwahl? 	<ul style="list-style-type: none"> ➤ N = 32 Lernende im Sekundarbereich im Fach Biologie ➤ Pretest (MC-Aufgaben verschiedener Komplexitätsniveaus zu Vererbungslehre) 	<ul style="list-style-type: none"> ➤ Post-Test-Ergebnisse für Einteilung in effektive (M = 4.35) und ineffektive Lerner (M = 2.27) Lerner; keine Unterschiede im Pretest zwischen den Gruppen ➤ Akkuratheit in der Gruppe der effektiven Lerner > als

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
			<ul style="list-style-type: none"> ➤ Es wird erwartet, dass effektive Lerner über eine bessere Akkuratheit der Selbsteinschätzung verfügen ➤ Es wird erwartet, dass effektive Lerner andere Kriterien bei der Aufgabenwahl berücksichtigen als ineffektive ➤ Es wird erwartet, dass effektive Lerner mehr Gebrauch von der Möglichkeit zur Selbsteinschätzung machen 	<ul style="list-style-type: none"> ➤ LCI: E-Learning bzgl. verschiedener Aufgaben (Formate s. Kostons, van Gog & Paas (2012)) ➤ Selbsteinschätzung nach jeder Aufgabe zu 7 Kriterien (Lösung, Ansatz, Zeitaufwand, Freude, Schwierigkeit, mentaler Aufwand, Gesamteinschätzung) auf neunstufiger Skala ➤ Nach Selbsteinschätzung Übersicht über alle Aufgaben und Auswahl ➤ Think-aloud während Selbsteinschätzung und Aufgabenwahl → Kodierung auf Basis von fünf Kategorien (Schlussfolgerungen, Vergleiche, Erklärungen, Erwähnungen, keine dergleichen) ➤ Abschließend Ranking der Wichtigkeit der Kriterien ➤ Post-Test unmittelbar nach LCI-Phase 	<p>in Gruppe der ineffektiven Lerner</p> <ul style="list-style-type: none"> ➤ Innerhalb der beiden Gruppen wiederum große Varianz in der Selbsteinschätzung → Bildung zweier Subgruppen ➤ Ineffektive Lerner nehmen mehr Vergleiche mit bisheriger Performanz vor, um sich selbst einzuschätzen ➤ Ineffektive Lerner nutzen Kriterium „Freude“ öfter zur Selbsteinschätzung ➤ Prädiktoren für schwierigere Aufgabenwahl: für ineffektive, akkurate Lernende (Lösung (-), Zeit (-), Gesamteinschätzung signifikante Parameter (+)); für effektive, akkurate Lernende (Zeit (-), Schwierigkeit (-), mentaler Aufwand (-))
8	Kostons, van Gog, & Paas (2012)	<ul style="list-style-type: none"> ➤ Kontrolle über die Wahl der zu bearbeitenden Aufgaben durch den Lernenden unterstützt Selbstregulation und Lernen; das gilt jedoch bedingt für Novizen (z. B. Azevedo, Moos, Greene, Winters, & Cromley, 2008; Goforth, 1994) 	<ul style="list-style-type: none"> ➤ Ist die Fähigkeit zur akkuraten Selbsteinschätzung und der Wahl adäquater Lernaufgaben trainierbar? ➤ Wie wirkt sich dies auf die Fähigkeit der Wahl 	<p>Experiment 1</p> <ul style="list-style-type: none"> ➤ N=80 Schüler im drittletzten Schuljahr ➤ 2x2-Design: Self-Assessment-Modell (ja vs. nein) & Task-Selection-Modell (ja vs. nein) → TG1: SA + TS, TG2: SA, 	<p>Experiment 1</p> <ul style="list-style-type: none"> ➤ Die Modellierung des Prozesses der Selbsteinschätzung führt zu signifikant höherer Akkuratheit ➤ Die Modellierung der Aufgabenauswahl hat keinen signifikanten Effekt auf die

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
		<ul style="list-style-type: none"> ➤ Cognitive Load als Erklärungsansatz dafür, dass Novizen Schwierigkeiten haben, eine akkurate Selbsteinschätzung zu geben (Sweller, Merrienboer, & Paas, 1998) ➤ Zone of proximal development (Vygotsky, 1978) 	<p>adäquater Lernaufgaben aus?</p> <ul style="list-style-type: none"> ➤ Welche Wirkungen zeigen sich hinsichtlich der Performanz? 	<p>TG3: TS, KG: weder SA noch TS</p> <ul style="list-style-type: none"> ➤ Pre- und Post-Test: 5 Aufgaben in 5 Niveaustufen zu Mendelschen Gesetzen (5 Lösungsschritte) ➤ Mentaler Aufwand auf neunstufiger Skala ➤ Selbsteinschätzung auf fünfstufiger Skala (ein Punkt pro Lösungsschritt) ➤ Nach Selbsteinschätzung Information über Schwierigkeit der gelösten Aufgabe, Auswahl weiterer Aufgabe (5 Komplexitätsstufen, 3 Unterstützungsstufen) 	<p>Akkuratheit der Selbsteinschätzung</p> <ul style="list-style-type: none"> ➤ Modellierung des Prozesses der Aufgabenauswahl fördert akkurate Auswahl von Lernaufgaben
				<p>Experiment 2</p> <ul style="list-style-type: none"> ➤ N=90 Schüler ➤ TG1: SA + TS (model), TG2: SA + TS (Erklärung der Self-Assessment-Kriterien & Übung), KG: keine Übung/Modellierung von SA + TS ➤ Vorgehensweise analog zu Experiment 1 ➤ Selbstregulierte Lernphase nach Training und vor Post-Test (acht Übungsaufgaben) 	<p>Experiment 2</p> <ul style="list-style-type: none"> ➤ Akkuratheit in TG1 > KG, jedoch keine Unterschiede zwischen TG1 und TG2 ➤ Post-Test zeigt keine signifikanten Unterschiede zwischen Gruppen bei adäquater Aufgabenauswahl ➤ In selbstregulierter Lernphase war adäquate Aufgabenauswahl in TG1 und TG2 besser als in KG ➤ Lernfortschritte für KG < als für TG1 und TG2 → höhere Akkuratheit führt zu besserer Selbstregulation

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					und letztlich besserer Performanz
9	Maki, Jonas, & Kallod (1994)	<ul style="list-style-type: none"> ➤ Rolle des Monitorings in Theorien des Textverständnisses (Ruddel & Speaker, 1985) ➤ Ausgangspunkt ist die indifferente Studienlage zum Zusammenhang zwischen dem Verständnis beim Lesen von Texten und der Fähigkeit der Vorhersage der Performanz 	<ul style="list-style-type: none"> ➤ Welche Zusammenhänge bestehen zwischen verschiedenen Instrumenten zur Messung des Textleseverständnisses und der Fähigkeit, die Performanz vorherzusagen bzw. nach einer Testung einzuschätzen? 	<ul style="list-style-type: none"> ➤ N = 75 Studierende der Psychologie ➤ Bearbeitung von Texten unter Einsatz der Multi-Media Comprehension Battery (MMCB) (Gernsbacher & Varner, 1988) und des Nelson-Denny Reading Test (Brown, Fishco, & Hanna, 1993) ➤ Vorhersage der Performanz auf sechsstufiger Skala ➤ Confidence Judgements nach Bearbeitung des Tests auf sechsstufiger Skala 	<ul style="list-style-type: none"> ➤ Vorhersage der Performanz inakkurat ($\gamma = .114$) ➤ Höhere Akkuratheit bei Confidence Judgements ($\gamma = .551$) ➤ Fast durchgängig signifikante Korrelationen zwischen Confidence Judgements und Testergebnissen (Performanz) → Probanden mit besserer Performanz weisen bessere Akkuratheit in der Selbsteinschätzung auf ➤ Zusammenhänge lassen sich nicht für Vorhersage der Performanz finden
10	Marshall & Brown (2006)	<ul style="list-style-type: none"> ➤ Positive affectivity vs. negative affectivity (Watson & Clark, 1984) 	<ul style="list-style-type: none"> ➤ Welche Zusammenhänge bestehen zwischen geringen Erwartungen hinsichtlich der Performanz einer Aufgabe und dem emotionalen Wohlbefinden des Lernenden? 	Studie 1 <ul style="list-style-type: none"> ➤ N = 81 Studierende ➤ Lernende bearbeiten computergestützten Assoziationstest (sinngemäße Ergänzung von Wortreihen, insgesamt zehn solcher Aufgaben, Zeitlimit von fünf Minuten) ➤ Eine Hälfte der Lernenden erhielt schwere Aufgaben, die andere Hälfte leichte 	Studie 1 <ul style="list-style-type: none"> ➤ Durchschnittliche Erwartung hinsichtlich der Zahl richtig gelöster Aufgaben von 5.2; Performanz für Gruppe mit leichteren Aufgaben höher ➤ Hohe Erwartung bzgl. der eigenen Performanz Prädiktor für Selbstwert ➤ Lernende mit hohen Erwartungen mit positiveren

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				<ul style="list-style-type: none"> ➤ Selbsteinschätzung/Erwartung auf neunstufiger Skala ➤ Zu Beginn und am Ende des Experiments Einsatz einer Skala zu Emotionen (4 Subskalen: Feeling of Self-worth, Happiness-Sadness, Agitation-Relaxation, Surprise) 	<p>Emotionen, unabhängig von Erfolg/Misserfolg</p> <ul style="list-style-type: none"> ➤ Zusammenhang mit allgemeinen emotionalen Tendenzen ➤ Lernende mit geringer erwarteter Performanz auch mit negativeren Emotionen → Überschätzung besser
		<ul style="list-style-type: none"> ➤ Attribuierungstheorie (Brown & Weiner, 1984; Weiner, 1986) 	<ul style="list-style-type: none"> ➤ Erwartung von Erfolg steht in Zusammenhang mit Performanz und sagt Fähigkeitsattribution voraus ➤ Fähigkeitsattribution steht in Zusammenhang mit Performanz und sagt Selbstwert voraus 	<p>Studie 2</p> <ul style="list-style-type: none"> ➤ N = 65 Studierende ➤ Design mit Studie 1 vergleichbar, bedeutsame Ausnahme: Selbsteinschätzung über Frage auf neunstufiger Skala, wie Performanz mit Fähigkeit zusammenhängt und wie akkurat der Test Fähigkeit misst 	<p>Studie 2</p> <ul style="list-style-type: none"> ➤ Attribuierung zu eigenen Fähigkeiten zeigt sich bei guter Performanz, unabhängig von der Erwartung
11	Mengelkamp & Bannert (2010)	<ul style="list-style-type: none"> ➤ Stabilität der Akkuratheit von Confidence Judgements ➤ Instrumente zur Bestimmung der Akkuratheit: absolute vs. relative Akkuratheit (Maki, 1998; Nelson, 1996) ➤ Ansatzpunkte, dass absolute Akkuratheit stabil ist, liefert Koriart (2007) mit Annahme der theory-based cues ➤ Akkurate Selbsteinschätzung als Prädiktor für die Performanz 	<ul style="list-style-type: none"> ➤ Inwiefern erweisen sich verschiedene Instrumente zur Erfassung der Akkuratheit von <i>confidence judgements</i> im Zeitverlauf als stabil? ➤ Inwiefern ist die Akkuratheit über zwei verschiedene Tests generalisierbar? ➤ Inwiefern sagt die Akkuratheit der Selbsteinschätzung voraus? 	<ul style="list-style-type: none"> ➤ N = 113 Studierende der Psychologie und Erziehungswissenschaften ➤ Virtuelle Lernumgebung (Hypermedia) mit verschiedenen Einträgen, Bildern & Links zum Thema „Operantes Konditionieren“ ➤ Einsatz von drei Wissenstests (Pre-test, intermediate Test nach zehn Minuten Lernen und abschließender Test nach weiteren 20 Minuten) mit verschiedenen kognitiven Anspruchsniveaus (Wissen, Verstehen, Transfer) 	<ul style="list-style-type: none"> ➤ Kurven zur Kalibrierung zeigen gering ausgeprägte Überschätzung für low-performer und eine geringe Unterschätzung für high-performer → insgesamt herrscht hohe Akkuratheit in der Selbsteinschätzung vor ➤ Stabilität der Selbsteinschätzung wurde für die Instrumente der absoluten

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
			schätzung die Performanz in einem abschließenden Test voraus?	<ul style="list-style-type: none"> ➤ Confidence Judgements für jede Aufgabe durch prozentuale Einschätzung der Richtigkeit der Antwort (25 %-Schritte bei MC-Aufgaben und kontinuierliche Skala bei offenen Aufgabenformaten) 	<p>Akkuratheit festgestellt, jedoch nicht für die relative Akkuratheit</p> <ul style="list-style-type: none"> ➤ Korrelationen zeigen tlw. Generalisierbarkeit der Fähigkeit zur Selbsteinschätzung, da Zusammenhänge zwischen verschiedenen Tests bestehen ➤ Prädiktive Validität für Instrumente der relativen Akkuratheit gegeben, weniger für absolute Akkuratheit der Selbsteinschätzung
12	Narciss, Koerndle, & Dresel (2011)	<ul style="list-style-type: none"> ➤ Nachteile der Überschätzung: Decision Affect Theory (Mellers, Schwartz, Ho, & Ritov, 1997), Kontrolltheorie der Selbstregulation (Carver, 2003; Carver & Scheier, 1990) & self-enhancement theory (Shrauger, 1975; Swann, Griffin, Predmore, & Gaines, 1987) ➤ Vorteile der Überschätzung: Selbstwirksamkeit (Bandura, 1977), Selbstkonzept gestärkt (Marsh, 1990; Marsh & O'Mara, 2008), günstigere Attribuierung von Erfolg (Brown & Marshall, 2001) 	<ul style="list-style-type: none"> ➤ Welche Auswirkungen haben Über-/Unterschätzung bzw. eine akkurate Selbsteinschätzung auf die Zufriedenheit mit der eigenen Performanz? ➤ Es wird angenommen, dass Unterschätzungen zu größerer Zufriedenheit führen, jedoch die Gefahr besteht, dass Erfolge „klein“ geredet werden. 	<ul style="list-style-type: none"> ➤ N = 67 Lernende der fünften Klasse (Gymnasium) ➤ Bildung von drei Gruppen: i) Kontrollgruppe (keine Übungen), ii) TG1 (Übung mit summativem Feedback zur Performanz), iii) TG2 (Übung mit Feedback zur Performanz; zusätzlich Hinweise Anstrengung/Attribuierung) ➤ Intervention besteht aus wöchentlichen Lerneinheiten à 45 Minuten, in denen Mathematikaufgaben geübt wurden ➤ Selbsteinschätzungen, Performanz und Zufriedenheit mit der Performanz wurden eine Woche vor und eine Woche nach der Übungsphase gemessen 	<ul style="list-style-type: none"> ➤ Treatment-Gruppen verbesserten Performanz signifikant, Kontrollgruppe nicht ➤ Fähigkeit zur akkuraten Selbsteinschätzung änderte sich im Zeitverlauf nicht, stärkste Ausprägung in der Kontrollgruppe ➤ TG2 mit geringerer Zufriedenheit bzgl. der Performanz ➤ Lernende, die ihre Performanz überschätzten, sind mit der Performanz zufriedener als jene, die sich unterschätzten → Nachteile von Unterschätzung überwiegen

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				<ul style="list-style-type: none"> ➤ Pre- und Post-Test: vier Rechentests mit je 15 Aufgaben ➤ Selbsteinschätzung auf Basis von nicht zu lösenden Beispielaufgaben (~ ease of learning judgements) ➤ Zufriedenheit mit der Performanz auf bipolarer zehnstufiger Skala 	
13	Nietfeld, Cao, & Osborne (2006)	<ul style="list-style-type: none"> ➤ (akkurates) Monitoring als Faktor, der Selbstregulation und Performanz begünstigt (Butler & Winne, 1995) ➤ Messung absoluter vs. relativer Akkuratheit der Selbsteinschätzung ➤ Kalibrierung als Prozess der Passung der Wahrnehmung der Performanz und tatsächlicher Ausprägung ➤ Selbstwirksamkeit 	<ul style="list-style-type: none"> ➤ Verbessern Übungen und Feedback zum Monitoring die Kalibrierung? ➤ Verbessern Übungen und Feedback zum Monitoring die Performanz? ➤ Ist die (angepasste) Kalibrierung verantwortlich für Änderungen in der Performanz/der Selbstwirksamkeit? 	<ul style="list-style-type: none"> ➤ N=84 Studierende der Pädagogischen Psychologie ➤ TG vervollständigt kurze Monitoring-Übungen am Ende einer Unterrichtseinheit (beinhalten auch Wiederholungsaufgaben mit Selbsteinschätzung der akkuraten Lösung auf kontinuierlicher Skala), Erhalt von Feedback zur Akkuratheit der Selbsteinschätzung ➤ KG gibt ebenfalls Selbsteinschätzung ab, erhält aber keine Übungen und Feedback ➤ Messung von Monitoring Accuracy (Kalibrierung = absolute Differenz zwischen Selbsteinschätzung und Performanz im Durchschnitt; bias score = Differenz zwischen durchschnittlicher Selbsteinschätzung und durchschnittlicher Performanz), Performanz (insgesamt 4 Tests mit je 20 MC-Aufgaben; Schema Repräsentation (Externalisierung eines mentalen Modells zu effektiver Instruktion) und Selbstwirksamkeit 	<ul style="list-style-type: none"> ➤ TG mit besserer Kalibrierung als KG im Zeitverlauf → Einfluss der Intervention ➤ Performanz in der TG ebenfalls signifikant > als jene in der KG ➤ Multiple Regressionsanalysen zeigen, dass Lernende, die Kalibrierung im Verlauf verbesserten, die Tendenz aufweisen, die Performanz ebenfalls zu verbessern → Kalibrierung kann für Performanzverbesserung verantwortlich gemacht werden ➤ Durchschnittliche Kalibrierung steht im Zusammenhang mit Änderungen der Selbstwirksamkeit, nicht jedoch die Anpassung der Kalibrierung

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
				(educational psychology self-efficacy inventory) ➤ Pretest (Vorwissenstest)	➤ Monitoring-Fähigkeiten verbessern auch Performanz in authentischen Aufgabensituationen (Schema Repräsentation)
14	Olina & Sullivan (2004)	<ul style="list-style-type: none"> ➤ Formatives Assessment (Sadler, 1989) ➤ Selbsteinschätzung als effektive Strategie selbstregulierten Lernens ➤ Metakognitiver Ansatz (Bransford, Brown, & Cocking, 1999) → Vorhersage der eigenen Performanz (Fokus auf Monitoring) 	<ul style="list-style-type: none"> ➤ Welche Effekte haben Selbsteinschätzung und Lehrerbewertung auf die Performanz der Lernenden? ➤ Verbessert die Kombination aus Lehrerbewertung und Selbsteinschätzung die Performanz der Lernenden überproportional? ➤ Haben die beiden Verfahren unterschiedliche Effekte auf die Einstellungen der Lernenden? 	<ul style="list-style-type: none"> ➤ N = 341 Lernende 10. und 11. Klassen ➤ Lernende durchlaufen Lerneinheit von 12 Unterrichtsstunden zur Einführung in die empirische Forschung, in denen Lernende selbst Studien zu Lernen durchführen und am Ende einen Projektbericht schreiben ➤ 4 TG/KG: keine Bewertung im Lernprozess (weder Feedback noch Selbsteinschätzung), Selbsteinschätzung (zwei im Laufe der Einheit), Fremdeinschätzung durch Lehrperson (zwei im Laufe der Einheit), Selbst- und Fremdeinschätzung (zwei im Laufe der Einheit) ➤ Lehrpersonen erhalten ausgearbeitete Instruktionsmaterialien ➤ Instrumente: Project Rating Scale (15 Items, dreistufige Skala, zur Selbsteinschätzung und Fremdeinschätzung eingesetzt), Posttest (21 MC- und Kurzantwortfragen zur Erfassung der Performanz), Erfassung der Einstellungen der Schüler und Lehrpersonen 	<ul style="list-style-type: none"> ➤ Höchste Performanzwerte bzgl. Projektbericht für TG Selbst- und Fremdeinschätzung, danach Fremdeinschätzung, Selbsteinschätzung & Kontrollgruppe ➤ Keine signifikanten Unterschiede im Posttest ➤ TG mit Selbsteinschätzung weisen höchste Einstellungswerte (Zufriedenheit, Überzeugung, Lerngewinn) auf ➤ Lernende sehen höchsten Nutzen in Fremd- und Selbsteinschätzung ➤ In beiden TG mit Selbsteinschätzung neigten Lernende zur Überschätzung (26.69 bzw. 27.36 vs. 18.85 bzw. 21.36 auf einer Skala bis 30)

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
15	Panadero & Romero (2014)	<ul style="list-style-type: none"> ➤ Monitoring als zentrale Voraussetzung für autonomes und selbstreguliertes Lernen (Zimmerman & Moylan, 2009) ➤ Self-Assessment = Einbindung des Lernenden in den Beurteilungsprozess (Falchikov & Boud, 1989) ➤ Verschiedene Ansätze zur Förderung der Fähigkeiten zur Selbsteinschätzung (Rubrics, Scripts, Prompts, Cues) 	<ul style="list-style-type: none"> ➤ Ziel ist der Vergleich der Effekte des Self-Assessments mit und ohne Bekanntgabe der Assessment-Kriterien (Rubrics) ➤ Wie beeinflussen Rubrics i. Vgl. zur einfachen Aufforderung zum Self-Assessment Selbstregulation, Performanz, Akkuratheit der Selbsteinschätzung und aufgabenbezogenen Stress? ➤ Es wird erwartet, dass der Einsatz von Rubrics zu höherer Selbstregulation, besserer Performanz, besserer Akkuratheit der Selbsteinschätzung, und geringerem aufgabenbezogenem Stress führt 	<ul style="list-style-type: none"> ➤ N = 218 angehende Lehrpersonen ➤ TG: Erhalt von Rubrics (n = 111), von Experten entwickelt; KG: kein Erhalt von Rubrics, aber Selbsteinschätzung (n = 107) ➤ 2x2-Design (rubrics vs. non-rubric & pre- und post-test) ➤ Einsatz des Emotion and Motivation Self-Regulation Questionnaire von Alonso Tapia et al. (2014) (20 Items, fünfstufige Skala, 2 Skalen: Learning self-regulation & performance/avoidance self-regulation) ➤ Einsatz offener Fragen zur Selbstregulation: Welche Strategien wurden zur Aufgabenbewältigung eingesetzt? ➤ Aufgabenbezogener Stress auf fünfstufiger Skala über ein Item erfasst 	<ul style="list-style-type: none"> ➤ TG mit höheren Werten für Learning Self-Regulation, aber auch für „negative“ Selbstregulation, gemessen durch performance/avoidance self-regulation als KG ➤ TG mit besseren Werten in der Performanz und in der Akkuratheit der Selbsteinschätzung als KG ➤ Aufgabenbezogener Stress stärker für TG ausgeprägt als für die KG
16	Sitzmann & Johnson, 2012	<ul style="list-style-type: none"> ➤ Feedback-Theorie (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991) ➤ Kontrolltheorie (Carver & Scheier, 1981) ➤ Einschätzungsvarianten: Über-/Unterschätzung, generell positiv/negativ 	Studie 1 <ul style="list-style-type: none"> ➤ Welche Auswirkungen hat eine inakurate Selbsteinschätzung auf die Performanz und das Aufsuchen von Lerngelegenheiten im weitergehenden Lernprozess? 	Studie 1 <ul style="list-style-type: none"> ➤ N = 293 Lernende (Großteil Beschäftigte) in einem Excel-Kurs (vierstündiger Online-Kurs, sechs Module) ➤ Freie Zeiteinteilung für Lernende im Kurs 	Studie 1 <ul style="list-style-type: none"> ➤ Selbsteinschätzung und Performanz korrelieren signifikant ➤ Aussteiger mit höherer Selbsteinschätzung

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
			<ul style="list-style-type: none"> ➤ H₁: Performanz für „generelle positiv-Gruppe“ am höchsten und Überschätzung am geringsten. ➤ H₂: Wahrscheinlichkeit des Ausscheidens für Überschätzung am höchsten und für „generell-positiv-Gruppe“ am niedrigsten. ➤ H₃: Der Zusammenhang von Selbsteinschätzung und Anstrengung mediert den Effekt, den die Selbsteinschätzung auf die Performanz hat; Anstrengung hat einen größeren positiven Effekt auf Performanz, wenn dem eine geringe Selbsteinschätzung vorangeht 	<ul style="list-style-type: none"> ➤ MC-Test (12 Items, deklaratives und prozedurales Wissen benötigt), Selbsteinschätzung (Skala von vier bis sechs Items, abhängig von Aufgabenzahl, Vorhersage der Zahl richtiger Antworten und fünfstufiger Skala, Konvertierung auf 100-Punkte-Skala) nach jedem Modul ➤ Erfassung des Ausstiegs nach einem Modul ➤ Anstrengung erfasst über Zeit, die Lernende im Lernsystem arbeiteten 	<ul style="list-style-type: none"> ➤ Selbsteinschätzung fällt von Modul 1 bis 5 und steigt zum Abschlussmodul ➤ Performanz sinkt bis Modul 4 und steigt danach bis zum Abschlussmodul ➤ Hoher Drop-Out; lediglich 24 % beendeten den Kurs ➤ Performanz am höchsten, wenn zuvor schon hoch und hohe Selbsteinschätzung (uniformly positive rating), gefolgt von Unterschätzung, Überschätzung und uniformly negative rating ➤ Analog dazu sind die Ergebnisse zum Kursausstieg → höchster Ausstieg bei Überschätzung, gefolgt von uniformly negative rating, Unterschätzung und uniformly positive rating ➤ Anstrengung hat einen größeren Einfluss auf den Effekt der Selbsteinschätzung auf die Performanz für niedrig ausgeprägte Selbsteinschätzungen als für hohe Selbsteinschätzung im vorangegangenen Modul; bei hoher Selbsteinschätzung

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					performen Lernende unabhängig von der Anstrengung
			Studie 2 <ul style="list-style-type: none"> ➤ Erweiterung des Modells um die Variablen Gewissenhaftigkeit (Barrick, Mount, & Strauss, 1993) und Einstellung ➤ Unabhängig von der Ausprägung der Gewissenhaftigkeit wird die Performanz im Folgemodul hoch sein für uniformly positive ratings und niedrig für uniformly negative ratings ➤ Analoge Ergebnisse für den Ausstieg aus dem Excelkurs erwartet ➤ Einstellung hat einen größeren Einfluss auf den Kursausstieg für Lernende mit geringen Selbsteinschätzungen 	Studie 2 <ul style="list-style-type: none"> ➤ N = 253 Lernende im Excel-Kurs ➤ Messung der Gewissenhaftigkeit vor dem Training (Mini-markers scale von Saucier), nach Beendigung eines Moduls Selbsteinschätzung (vier Items; Vorhersage richtiger Antworten) und Performanztestung (6 MC-Aufgaben) ➤ Einstellung über drei Items erfasst ➤ Vorwissen kontrolliert 	Studie 2 <ul style="list-style-type: none"> ➤ Selbsteinschätzung steht in Zusammenhang mit Performanz ➤ Unabhängig von der Gewissenhaftigkeit ist die Performanz im Folgemodul für uniformly positive ratings am höchsten und für uniformly negative ratings am niedrigsten ➤ Performanz im Folgemodul sinkt für Lernende mit hoher Ausprägung der Gewissenhaftigkeit ➤ Unabhängig von der Gewissenhaftigkeit ist der Ausstieg für uniformly positive ratings gering und für uniformly negative ratings hoch ➤ Unter- und Überschätzung führen für Lernende mit hoher Gewissenhaftigkeit zu weniger Ausstiegen aus dem Kurs ➤ Die Einstellung zum Training hat einen stärkeren Ef-

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					fekt auf die Wahrscheinlichkeit des Ausstiegs, wenn Selbsteinschätzungen gering ausgeprägt sind
17	Snyder, Nietfeld, & Linnenbrink-Garcia (2011)	<ul style="list-style-type: none"> ➤ Metakognition = Wissen über Kognition (Baker, 1989; Gregory Schraw, 1998), d.h. Wissen über kognitive Prozesse, Strategien und deren Einsatz und die eigenen kognitiven Fähigkeiten sowie Regulation kognitiver Prozesse (Flavell, 1979; Schraw, 1998), d.h. aktives Monitoring der kognitiven Prozesse und der Nutzung der eingesetzten Strategien ➤ Kalibrierung = Passung zwischen Wahrnehmung der eigenen Performanz und der tatsächlichen Performanz (Dunlosky & Metcalfe, 2009) ➤ Metakognition als mögliches Begründungsmuster für besseres Abschneiden von begabten Lernenden 	<ul style="list-style-type: none"> ➤ Welche Unterschiede bestehen in der Metakognition (confidence judgements global und local) zwischen begabten und durchschnittlichen Lernenden? ➤ Es wird erwartet, dass begabte Lernende über bessere Fähigkeit zur Kalibrierung, kurzfristig bessere Akkuratheitsentwicklung, bessere Vorhersagen für Examen und schließlich besserer Performanz verfügen 	<ul style="list-style-type: none"> ➤ N = 67 Schüler der zehnten Klasse in Biologie, homogene Gruppen hochbegabter und normal begabter Lernender ➤ Messung der Metakognition durch Metacognitive Awareness Inventory (MAI, 52 Items zu knowledge of cognition & regulation of cognition, Schraw & Dennison, 1994) ➤ Vorhersage der Anzahl richtiger Antworten (pre- & postdiction; global judgement), Differenz zwischen Performanz und Einschätzung = global judgement accuracy ➤ Confidence judgements auf kontinuierlicher Skala (0-100) für alle Items, Bildung der durchschnittlichen Differenz aus Einschätzung und tatsächlicher Performanz als Index der Akkuratheit, Bias der Kalibrierung als Differenz zwischen durchschnittlicher Selbsteinschätzung und durchschnittliche Performanz ($< 0 \rightarrow$ Unterschätzung) ➤ Performanzmessung via 4 Prüfungen 	<ul style="list-style-type: none"> ➤ Keine Unterschiede hinsichtlich des MAI zwischen hochbegabten und normalbegabten Lernenden ➤ Hochbegabte Lerner übertreffen normalbegabte in der Performanz über alle Prüfungen hinweg ➤ Hochbegabte mit akkuraterem Monitoring (local calibration) ➤ Bezüglich globaler Vorhersage der Anzahl der richtigen Antworten keine Unterschiede, bei der retrospektiven Einschätzung weisen hochbegabte akkuratere Werte auf ➤ Keine Unterschiede im Bias der Kalibrierung ➤ Multiple Regressionsanalysen zeigen Zusammenhänge zwischen Genauigkeit der local calibration und der Performanz \rightarrow Fähigkeit der Hochbegabten zum akkuratem Monitoring

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					unterstützt diese in der Performanz
18	Thiede, Anderson, & Theriault (2003)	<ul style="list-style-type: none"> ➤ Selbstreguliertes Lernen als Diskrepanz-Reduktions-Modell (Butler & Winne, 1995; Dunlosky & Hertzog, 2005): Lernen mit fortlaufendem Abgleich von IST- und SOLL-Zustand erfordert akkurates Monitoring ➤ Allerdings kaum empirische Beweise für den Zusammenhang von akkuratem Monitoring und Lernen 	<ul style="list-style-type: none"> ➤ Welcher Zusammenhang besteht zwischen der Akkuratheit des Monitorings, der Selbstregulation und der Performanz? ➤ Es wird erwartet, dass eine höhere Akkuratheit eine effektivere Regulation des Lernens bewirkt und somit zu besseren Lernergebnissen führt. 	<ul style="list-style-type: none"> ➤ N=66 Studierende der (Pädagogischen) Psychologie ➤ 3 Gruppen: TG1 (Lesen von sechs Texten und nach Lesen aller Texte Generieren von sechs Schlüsselbegriffen), TG2 (Lesen der sechs Texte, nach Lesen eines jeden Textes Generieren von Schlüsselbegriffen), KG (Lesen der sechs Texte) ➤ Ease of Learning-Einschätzung auf siebenstufiger Skala ➤ Anschließend Selbsteinschätzung des Verständnisses der Texte für alle Gruppen (siebenstufige Skala), Test, Feedback zu allgemeinem Abschneiden (nicht textspezifisch), Wiederholung ausgewählter Texte, finaler Test 	<ul style="list-style-type: none"> ➤ Akkuratheit mittels Goodman-Kruskal-Korrelation ermittelt: Akkuratheit unterscheidet sich zwischen den Gruppen (höchste Akkuratheit bei TG1, gefolgt von KG und TG2) ➤ Lernende aus TG1 wiederholten eher nicht verstandene Texte als TG2 und KG ➤ Performanz von TG1 übertrifft TG2 und KG im zweiten Test
19	van Loon, de Bruin, van Gog, & van Merriënboer (2013)	<ul style="list-style-type: none"> ➤ Judgements of Learning (relative accuracy) vs. Self-Score-Judgements (absolute accuracy/calibration, (Maki, Shields, Wheeler, & Zacchilli, 2005)) ➤ Cue utilization Framework von Koriat (1997) als Erklärungsmodell für inakkurate Selbsteinschätzungen: Urteile basieren auf falschen, 	<ul style="list-style-type: none"> ➤ Inwiefern führt die Aktivierung von inakkuratem Vorwissen eher zur Überschätzung als der Verzicht der Aktivierung des Vorwissens? ➤ Es wird erwartet, dass die Aktivierung von inakkuratem Vorwissen zu geringerer Perfor- 	Studie 1 <ul style="list-style-type: none"> ➤ N = 103 Primarschüler der 3. & 4. Klasse ➤ Pretest über zu lernende Konzepte, Lernen von 20 neuen Konzepten, Selbsteinschätzung auf Basis elfstufiger Skala für jedes gelernte Konzept (JoL), Auswahl von Konzepten zum wiederholten Lernen (kein er- 	Studie 1 <ul style="list-style-type: none"> ➤ Relative Akkuratheit (JoL) ist hoch (γ-Korrelation .78), starker Zusammenhang mit Wahl der wiederholt zu lernenden Inhalte (γ-Korrelation -.80) ➤ Lernende weisen eine hohe Überschätzung auf (Self-Score Judgements), wenn

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
		<p>für die Performanz nicht maßgeblichen Kriterien</p> <ul style="list-style-type: none"> ➤ Verfügbarkeit von akkuratem Vorwissen als zentraler Baustein für die Selbsteinschätzung → inakkurates Vorwissen nachteilig für Monitoring 	manz, zu mehr Überschätzung und weniger Selbstregulation führt	neutes Lernen), Posttest mit gleichen Fragen wie im Pretest, Selbsteinschätzung (Self-Score-Judgement auf dreistufiger Skala)	<p>sie fehlerhafte Aufgaben selbst einschätzen, die Überschätzung ist höher, wenn zuvor inakkurates Vorwissen aktiviert worden ist</p> <ul style="list-style-type: none"> ➤ Wahl geeigneter Items zur Wiederholung ist für Probanden mit falschem aktiviertem Vorwissen weniger adäquat als für eine fehlende Aktivierung von Vorwissen
				<p>Studie 2</p> <ul style="list-style-type: none"> ➤ N = 95 Primarschüler der 5. & 6. Klasse ➤ Methodisches Vorgehen entspricht jener aus Studie 1, lediglich zu lernende Konzepte wurden geändert 	<p>Studie 2</p> <ul style="list-style-type: none"> ➤ Relative Akkuratheit (JoL) ist hoch (γ-Korrelation .76), starker Zusammenhang mit Wahl der wiederholt zu lernenden Inhalte (γ-Korrelation -.81) ➤ Lernende weisen Überschätzung auf (Self-Score Judgements) ➤ Für falsch wiedergegebene Items ist Selbsteinschätzung (JoL) höher bei inakkurat aktiviertem Vorwissen als bei keiner Aktivierung des Vorwissens, gleiches gilt für die Überschätzung im Rahmen der Self-Score Judgements

Nr.	Autoren	Theoretischer Bezugsrahmen	Fragestellungen/ Hypothesen	Design	Ergebnisse
					<ul style="list-style-type: none">➤ Wahl geeigneter Items zur Wiederholung ist für Probanden mit falsch aktiviertem Vorwissen weniger adäquat als für eine fehlende Aktivierung von Vorwissen➤ Befunde aus Studie 1 bestätigt, etwas weniger starke Ausprägungen bei der Überschätzung

Anhang A2 Codebuch

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
EZP EZP		Erhebungszeitpunkt	1 = Übung Mittwoch 2 = Übung Donnerstag
IDF		Fragebogennummer	
Code1 Code2	Die Summe der <u>ersten drei Ziffern</u> der Postleitzahl des Heimatortes vor Studienbeginn (z. B. <u>33442</u> = 10)... Die Summe aus <u>Geburts</u> tag und <u>Geburts</u> monat Ihrer Mutter (z. B. <u>26.02.</u> = 28)... Die Summe der <u>letzten drei Ziffern</u> Ihrer Matrikelnummer (z. B. <u>1086325</u> = 10)...	Identifikationscode Fragebogen	
SE121 SE221	Geben Sie zunächst an, welche Aussage Ihren derzeitigen Wissens- und Könnensstand am ehesten widerspiegelt. Bitte kreuzen Sie eine Antwortalternative an.	Einschätzung des Wissens- und Könnensstands bei Auswahl einer Alternative vor der Aufgabenbearbeitung	1 = Mir fehlt noch das Verständnis für die Grundlagen der Buchführung. 2 = Ich bin in der Lage, grundlegende Begriffe und Prinzipien der Buchführung unabhängig voneinander in wenigen Sätzen wiederzugeben (z. B. Definition erfolgswirksamer Geschäftsvorfälle). 3 = Ich bin in der Lage, die Zusammenhänge zwischen grundlegenden Begriffen und Prinzipien der Buchführung zu erläutern (z. B. Bestimmung erfolgswirksamer Geschäftsvorfälle). 4 = Ich bin in der Lage, Buchungssätze zu gegebenen Geschäftsvorfällen aufzustellen.

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
			5 = Ich habe grundlegende Begriffe und Prinzipien der Buchführung verstanden und bin in der Lage, sie auf neuartige Aufgabenstellungen anzuwenden. 7 = Mehrfachantwort 9 = fehlende Antwort
SE122 SE222	Mir fehlt noch das Verständnis für die Grundlagen der Buchführung.	fehlendes Verständnis vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE123 SE223	Ich bin in der Lage, grundlegende Begriffe und Prinzipien der Buchführung unabhängig voneinander in wenigen Sätzen wiederzugeben (z. B. Definition erfolgswirksamer Geschäftsvorfälle).	Wiedergabe grundlegender Begrifflichkeiten und Prinzipien vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE124 SE224	Ich bin in der Lage, die Zusammenhänge zwischen grundlegenden Begriffen und Prinzipien der Buchführung zu erläutern (z. B. Bestimmung erfolgswirksamer Geschäftsvorfälle).	Erläuterung von Zusammenhängen vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE125 SE225	Ich bin in der Lage, Buchungssätze zu gegebenen Geschäftsvorfällen aufzustellen.	Aufstellung von Buchungssätzen vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE126 SE226	Ich habe grundlegende Begriffe und Prinzipien der Buchführung verstanden und bin in der Lage, sie auf neuartige Aufgabenstellungen anzuwenden.	Transfer auf neue Aufgaben vor der Aufgabenbearbeitung (Initiierungszeitpunkt I)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
TA131 TA231	Was versteht man unter dem <i>Prinzip der getrennten Kontenführung</i> ?	Posten der Aktivseite dürfen nicht mit Posten der Passivseite, [...] Aufwendungen nicht mit Erträgen verrechnet werden. Der Sache nach gleiche Erfolgskomponenten werden auf getrennten Konten erfasst.	0 = vollständig fehlerhafte Antwort 1 = ein Aspekt genannt (Verrechnungsverbot ODER Erfassung gleicher Erfolgskomponenten auf getrennten Konten) 2 = vollständig korrekte Antwort (Verrechnungsverbot UND Erfassung gleicher Erfolgskomponenten auf getrennten Konten) 8 = Spaßantwort 9 = fehlende Antwort
SE131 SE231	Ich bin überzeugt, die Aufgabe richtig gelöst zu haben.	Überzeugung nach Bearbeitung der ersten Textaufgabe	1 = gar nicht überzeugt ... 6 = sehr überzeugt 9 = fehlende Antwort
TA132 TA232	Nennen Sie die in §1 UStG genannten Tatbestandselemente steuerbarer Umsätze.	Lieferung oder sonstige Leistung, Unternehmereigenschaft, Tätigwerden im Inland, Tätigwerden im Rahmen eines Unternehmens, Leistung gegen Entgelt	0 = kein Element korrekt genannt 1 = ein Element korrekt genannt ... 5 = fünf Elemente korrekt genannt 8 = Spaßantwort 9 = fehlende Antwort
SE132 SE232	Ich bin überzeugt, die Aufgabe richtig gelöst zu haben.	Überzeugung nach Bearbeitung der zweiten Textaufgabe	1 = gar nicht überzeugt ... 6 = sehr überzeugt 9 = fehlende Antwort
TA133 TA233	Was versteht man unter dem <i>Belegprinzip</i> ? Gehen Sie auf die Anforderungen einer ordnungsmäßigen Belegorganisation ein.	Es gilt der Grundsatz "keine Buchung ohne Beleg". Schriftstücke, die Richtigkeit der Angaben beweisen. Nachweisbarkeit von	0 = kein Aspekt korrekt genannt 1 = ein Aspekt korrekt genannt ... 4 = vier Aspekte korrekt genannt 8 = Spaßantwort

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
		Buchung und Beleg. Aufbewahrungsfrist von 10 Jahren.	9 = fehlende Antwort
SE133 SE233	Ich bin überzeugt, die Aufgabe richtig gelöst zu haben.	Überzeugung nach Bearbeitung der dritten Textaufgabe	1 = gar nicht überzeugt ... 6 = sehr überzeugt 9 = fehlende Antwort
MC141 MC241	Zinsertrag an Bank	Aufwand/Ertrag	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
MC142 MC242	Privatentnahme an Bank	Erfolgsneutrale Buchung	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
MC143 MC243	Rohstoffaufwand an Rohstoffe	Aufwand/Ertrag	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
MC144 MC244	GuV-Konto an Eigenkapital	Kontenabschluss	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
MC145 MC245	Kurzfristige Verbindlichkeiten an langfristige Verbindlichkeiten	Erfolgsneutrale Buchung, Passivtausch	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
MC146 MC246	Fuhrpark an Privateinlage	Erfolgsneutrale Buchung	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
MC147 MC247	Mietaufwand an Bank	Aufwand/Ertrag	0 = fehlerhafte Antwort 1 = korrekte Antwort 9 = fehlende Antwort
SE148 SE248	Ich bin überzeugt, die Aufgaben 4.1 bis 4.7 richtig gelöst zu haben.	Überzeugung nach Bearbeitung der MC-Aufgaben	1 = gar nicht überzeugt ... 6 = sehr überzeugt 9 = fehlende Antwort
SE151 SE251	Mir fehlt noch das Verständnis für die Grundlagen der Buchführung.	fehlendes Verständnis nach der Bearbeitung von Text- und MC-Aufgaben (Initiierungszeitpunkt II)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE152 SE252	Ich bin in der Lage, grundlegende Begriffe und Prinzipien der Buchführung unabhängig voneinander in wenigen Sätzen wiederzugeben (z. B. Definition erfolgswirksamer Geschäftsvorfälle).	Wiedergabe grundlegender Begrifflichkeiten und Prinzipien nach der Bearbeitung von Text- und MC-Aufgaben (Initiierungszeitpunkt II)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE153 SE253	Ich bin in der Lage, die Zusammenhänge zwischen grundlegenden Begriffen und Prinzipien der Buchführung zu erläutern (z. B. Bestimmung erfolgswirksamer Geschäftsvorfälle).	Erläuterung von Zusammenhängen nach der Bearbeitung von Text- und MC-Aufgaben (Initiierungszeitpunkt II)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE154 SE254	Ich bin in der Lage, Buchungssätze zu gegebenen Geschäftsvorfällen aufzustellen.	Aufstellung von Buchungssätzen nach der Bearbeitung von Text- und MC-Aufgaben (Initiierungszeitpunkt II)	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
SE155 SE255	Ich habe grundlegende Begriffe und Prinzipien der Buchführung verstanden und bin	Transfer auf neue Aufgaben nach der Bearbeitung von Text- und MC-Aufgaben	1 = gar nicht zutreffend ... 6 = sehr zutreffend

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
	in der Lage, sie auf neuartige Aufgabenstellungen anzuwenden.	(Initiierungszeitpunkt II)	9 = fehlende Antwort
BS161 BS261	Auf dem Bankkonto gehen 450 € für eine an den Unternehmer Z vermietete Lagerhalle ein.	Bank 450 € an Mietetrug 450 €	0 = fehlerhafter Buchungssatz 1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt
BS162 BS262	Tilgung eines Bankkredits: Neben der Tilgungsrate i. H. v. 3.000 € sind die fälligen Zinsen i. H. v. 120 € von der M-OHG zu überweisen.	Zinsaufwand 120 € Verbindlichkeiten 3.000 an Bank 3.120 €	0 = fehlerhafter Buchungssatz 1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt
BS163 BS263	Am Geschäftsjahresende ist bei der M-OHG das Konto „Privatentnahmekonto“ abzuschließen. Die Summe der Privatentnahmen der laufenden Geschäftsperiode beträgt insgesamt 25.900 €.	Eigenkapital 25.900 € an Privatentnahmekonto 25.900 €	0 = fehlerhafter Buchungssatz 1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt
BS164 BS264	Um den Ablauf beim Postversand zu beschleunigen, kauft die M-OHG eine Franchiermaschine für 260 € (netto) auf Ziel.	BGA 260 € Vorsteuer 26 € an Verbindlichkeiten 286 €	0 = fehlerhafter Buchungssatz 1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt
BS165 BS265	Unternehmer M legt 4.500 € in die betriebliche Kasse der M-OHG ein.	Kasse 4.500 € an Privateinlage 4.500 €	0 = fehlerhafter Buchungssatz

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
			1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt
BS166 BS266	Verkauf von Waren im Gesamtwert von 55.000 € (brutto) auf Ziel.	Forderungen aus LuL an UE HW 50.000 € USt. 5.000 €	0 = fehlerhafter Buchungssatz 1 = Buchungssatz tlw. korrekt (richtige Konten gewählt, jedoch keine/falsche Beträge) 2 = Buchungssatz korrekt aufgestellt (auch bei Berechnung der USt.-Last mit den gültigen USt.-Sätzen)
SE167 SE267	Ich bin überzeugt, die Aufgaben 6.1.bis 6.6 richtig gelöst zu haben.	Überzeugung nach Bearbeitung der Buchungsaufgaben	1 = gar nicht überzeugt ... 6 = sehr überzeugt 9 = fehlende Antwort
SE171 SE271	Ich bin in der Lage, Buchungssätze zu gegebenen Geschäftsvorfällen aufzustellen.	Aufstellung von Buchungssätzen nach der Bearbeitung der Buchungsaufgaben	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
FS181 FS281	Ermitteln Sie den Gewinn und das Eigenkapital der M-OHG für das Geschäftsjahr t_1 anhand der folgenden Angaben: Anlagevermögen am 31.12. t_1 : 350.000 € Umlaufvermögen am 31.12. t_1 : 250.000 € Verschuldungsgrad am 31.12. t_1 : 4	Geschäftsjahresgewinn = $EK_{\text{neu}} - EK_{\text{alt}} + \text{Privatentnahmen} - \text{Privateinlagen}$ AV + UV = EK + FK $EK_1 = 125.000 \text{ €}$	je ein Punkt für Formel zur Ermittlung des Geschäftsjahresgewinns sowie der Bilanzgleichung ein Punkt für Bestimmung des Eigenkapital _{neu} ein Punkt für Berechnung des Geschäftsjahresgewinns

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
	<p>Privatentnahmen während des Geschäftsjahres t_1: 2.500 € monatlich</p> <p>Privateinlage am 30.09.t_1: 20.000 € Privateinlage am 15.03.t_1: 5.000 € Eigenkapital am 31.12.t_0: 100.000 €</p> <p><i>Hinweis: der Verschuldungsgrad ist definiert als Fremdkapital/Eigenkapital</i></p>	<p>Geschäftsjahresgewinn = 125.000 € - 100.000 € + 12*2.500 € - 20.000 € - 5.000 € = 30.000 €</p>	
SE182 SE282	Ich bin überzeugt, die Aufgabe richtig gelöst zu haben.	Überzeugung nach Bearbeitung der Transferaufgabe	<p>1 = gar nicht überzeugt ...</p> <p>6 = sehr überzeugt</p> <p>9 = fehlende Antwort</p>
SE191 SE291	Ich habe grundlegende Begriffe und Prinzipien der Buchführung verstanden und bin in der Lage, sie auf neuartige Aufgabenstellungen anzuwenden.	Transfer auf neue Aufgaben nach der Bearbeitung der Transferaufgabe	<p>1 = gar nicht zutreffend ...</p> <p>6 = sehr zutreffend</p> <p>9 = fehlende Antwort</p>
LS1101 LS2101	Ich habe die Vorlesung regelmäßig besucht.	Besuch der Vorlesung	<p>1 = trifft gar nicht zu (> 2 Fehltermine)</p> <p>2 = trifft weitgehend zu (1-2 Fehltermine)</p> <p>3 = trifft genau zu (kein Fehltermin)</p> <p>9 = fehlende Antwort</p>
LS1102 LS2102	Ich habe die Übung regelmäßig besucht.	Besuch der Übung	<p>1 = trifft gar nicht zu (> 2 Fehltermine)</p> <p>2 = trifft weitgehend zu (1-2 Fehltermine)</p> <p>3 = trifft genau zu (kein Fehltermin)</p> <p>9 = fehlende Antwort</p>

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
LS1103 LS2103	Ich habe die Tutorien regelmäßig besucht.	Besuch der Tutorien	1 = trifft gar nicht zu (> 2 Fehltermine) 2 = trifft weitgehend zu (1-2 Fehltermine) 3 = trifft genau zu (kein Fehltermin) 9 = fehlende Antwort
LS1104 LS2104	Ich bereite die Inhalte der Vorlesung vor .	Vorbereitung der Vorlesung	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS1105 LS2105	Ich bereite die Inhalte der Übung vor .	Vorbereitung der Übung	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS1106 LS2106	Ich bereite die Inhalte der Tutorien vor .	Vorbereitung der Tutorien	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS1107 LS2107	Ich bereite die Inhalte Vorlesung nach .	Nachbereitung der Vorlesung	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS1108 LS2108	Ich bereite die Inhalte der Übung nach .	Nachbereitung der Übung	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS1109 LS2109	Ich bereite die Inhalte der Tutorien nach .	Nachbereitung der Tutorien	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
LS11010 LS21010	Ich versuche, die bereitgestellten Übungsaufgaben eigenständig zu lösen.	Übungsaufgaben eigenständig lösen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11011 LS21011	Ich bearbeite die bereitgestellten Übungsaufgaben gemeinsam mit Kommilitonen.	Übungsaufgaben mit Kommilitonen lösen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11012 LS21012	Bevor ich Übungsaufgaben löse, gehe ich die Vorlesungsunterlagen nochmals durch.	Vorlesungsunterlagen lesen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11013 LS21013	Wenn ich einfachere Übungsaufgabe sicher lösen kann, wende ich mich rasch schwierigeren Aufgaben zu.	rasch schwierige Übungsaufgaben lösen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11014 LS21014	Ich wiederhole oft einfache Übungsaufgaben, auch wenn ich sie schon mehrmals richtig gelöst habe.	einfache Übungsaufgaben wiederholen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11015 LS21015	Wenn ich bei einer Aufgabe nicht weiterkomme, stelle ich diese zunächst zurück, um sie später zu lösen.	schwierige Übungsaufgaben zurückstellen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11016 LS21016	Wenn ich bei einer Aufgabe nicht weiterkomme, frage ich Kommilitonen um Rat.	Kommilitonen um Rat fragen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
LS11017 LS21017	Wenn ich bei einer Aufgabe nicht weiterkomme, gehe ich die vorlesungsbegleitende Literatur noch einmal genau durch.	vorlesungsbegleitende Literatur lesen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
LS11018 LS21018	Wenn ich bei einer schwierigen Übungsaufgabe nicht weiterkomme, bearbeite ich zunächst einfachere Aufgaben.	einfache Übungsaufgaben zuerst lösen	1 = gar nicht... 6 = sehr regelmäßig 9 = keine Antwort
ZA1111 ZA2111	Ich habe im Rahmen der Veranstaltung "Grundlagen des externen Rechnungswesens" in den vergangenen Wochen im Vergleich zu anderen Veranstaltungen meines Studiums sehr viel mehr Zeit investiert.	relativer Zeitaufwand	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
ZA1112 ZA2112	Zeit für Vor- und Nachbereitung der Vorlesung insgesamt [Angabe in Minuten pro Woche]	absoluter Zeitaufwand Vorlesung [Minuten pro Woche]	
ZA1113 ZA2113	Zeit für Vor- und Nachbereitung der Übung insgesamt [Angabe in Minuten pro Woche]	absoluter Zeitaufwand Übung [Minuten pro Woche]	
ZA1114 ZA2114	Zeit für Vor- und Nachbereitung der Tutorien insgesamt [Angabe in Minuten pro Woche]	absoluter Zeitaufwand Tutorium [Minuten pro Woche]	
MO1121 MO2121	Ich erachte meine Motivation, neue Inhalte im Fach Rechnungswesen zu lernen, für hoch.	Motivation für neue Inhalte des Rechnungswesens	1 = gar nicht zutreffend ... 6 = sehr zutreffend 9 = fehlende Antwort
AA1131 AA2131	Sind Sie weiblich oder männlich?	Geschlecht	1 = weiblich 2 = männlich 9 = fehlende Antwort
AA1132	Wie alt sind Sie?	Alter	

Itembezeichnung	Itemformulierung	Itembeschreibung/ Korrekte Antwort	Werte
AA2132			
AA1133 AA2133	Welches Studienfach studieren Sie?	Studienfach	1 = Bachelor Kultur und Wirtschaft 2 = Bachelor Unternehmensjurist 3 = Bachelor Volkswirtschaftslehre 4 = Bachelor Wirtschaftsinformatik 5 = Bachelor Wirtschaftspädagogik 5 = Bachelor Wirtschaftsmathematik 6 = Lehramt an Gymnasien 6 = Bachelor Wirtschaftspädagogik 7 = Master Psychologie 7 = Lehramt an Gymnasien 8 = sonstiges 8 = Master Psychologie 9 = fehlende Antwort 9 = sonstiges
AA1134 AA2134	An welchem Schultyp haben Sie Ihre Hochschulreife erworben?	HZB	1 = allgemeinbildendes Gymnasium 2 = Wirtschaftsgymnasium 3 = anderer Schultyp 9 = fehlende Antwort
AA1135 AA2135	Haben Sie bereits eine Berufsausbildung im kaufmännischen Bereich abgeschlossen?	kaufmännische Berufsausbildung	1 = ja 2 = nein 9 = fehlende Antwort
AA1136 AA2136	Haben Sie sich bereits in der Schule häufig mit Inhalten des externen Rechnungswesens beschäftigt?	schulische Erfahrungen mit Rechnungswesen	1 = ja 2 = nein 9 = fehlende Antwort

Anhang A3 Mittelwertvergleich der Selbsteinschätzungen der Gesamtstichprobe

	Initiierungszeit- punkt I		Initiierungszeit- punkt II				
Messzeitpunkt I	(328 ≤ n ≤ 374)						
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
Fehlendes Verständnis	2.96	1.55	3.56	1.60	-9.25	374	.000
Wiedergabe grundlegen- der Begriffe und Prinzi- pien	3.50	1.33	2.79	1.24	11.30	372	.000
Erläuterung von Zusammenhängen	3.22	1.31	2.72	1.25	8.58	371	.000
Aufstellung von Buchungssätzen I	3.96	1.53	3.23	1.50	11.92	371	.000
Transfer auf neue Aufgaben I	2.79	1.34	2.43	1.29	6.49	370	.000
Aufstellung von Buchungssätzen II ⁴⁶	-	-	2.90	1.46	14.71	363	.000
Transfer auf neue Aufgaben II	-	-	2.16	1.21	9.35	328	.000
Messzeitpunkt II	(152 ≤ n ≤ 162)						
Fehlendes Verständnis	2.09	1.07	2.52	1.28	-4.58	160	.000
Wiedergabe grundlegen- der Begriffe und Prinzi- pien	4.08	1.13	3.50	1.13	-6.84	161	.000
Erläuterung von Zusammenhängen	4.04	1.07	3.59	1.18	5.56	161	.000
Aufstellung von Buchungssätzen I	4.70	1.08	4.36	1.10	4.08	161	.000
Transfer auf neue Aufgaben I	3.59	1.22	3.47	1.25	1.64	161	.103
Aufstellung von Buchungssätzen II	-	-	4.63	1.11	.84	157	.405
Transfer auf neue Aufgaben II	-	-	3.34	1.29	2.67	151	.008

⁴⁶ Die Mittelwertvergleiche der Aufstellung von Buchungssätzen II sowie des Transfers auf neue Aufgaben II wurden jeweils mit den Angaben zum ersten Initiierungszeitpunkt vorgenommen.

Anhang A4 Mittelwertvergleich der Selbsteinschätzungen vor der Aufgabenbearbeitung für beide Messzeitpunkte (N = 56)

	Messzeitpunkt I		Messzeitpunkt II		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Fehlendes Verständnis	2.80	1.29	1.96	1.03	-4.23	55	.000
Wiedergabe grundlegender Begriffe und Prinzipien	3.57	1.37	4.23	1.01	3.62	55	.001
Erläuterung von Zusammenhängen	3.30	1.26	4.16	1.09	4.83	55	.000
Aufstellung von Buchungssätzen	4.14	1.34	4.82	.86	3.76	55	.000
Transfer auf neue Aufgaben	2.82	1.21	3.70	1.08	5.80	55	.000

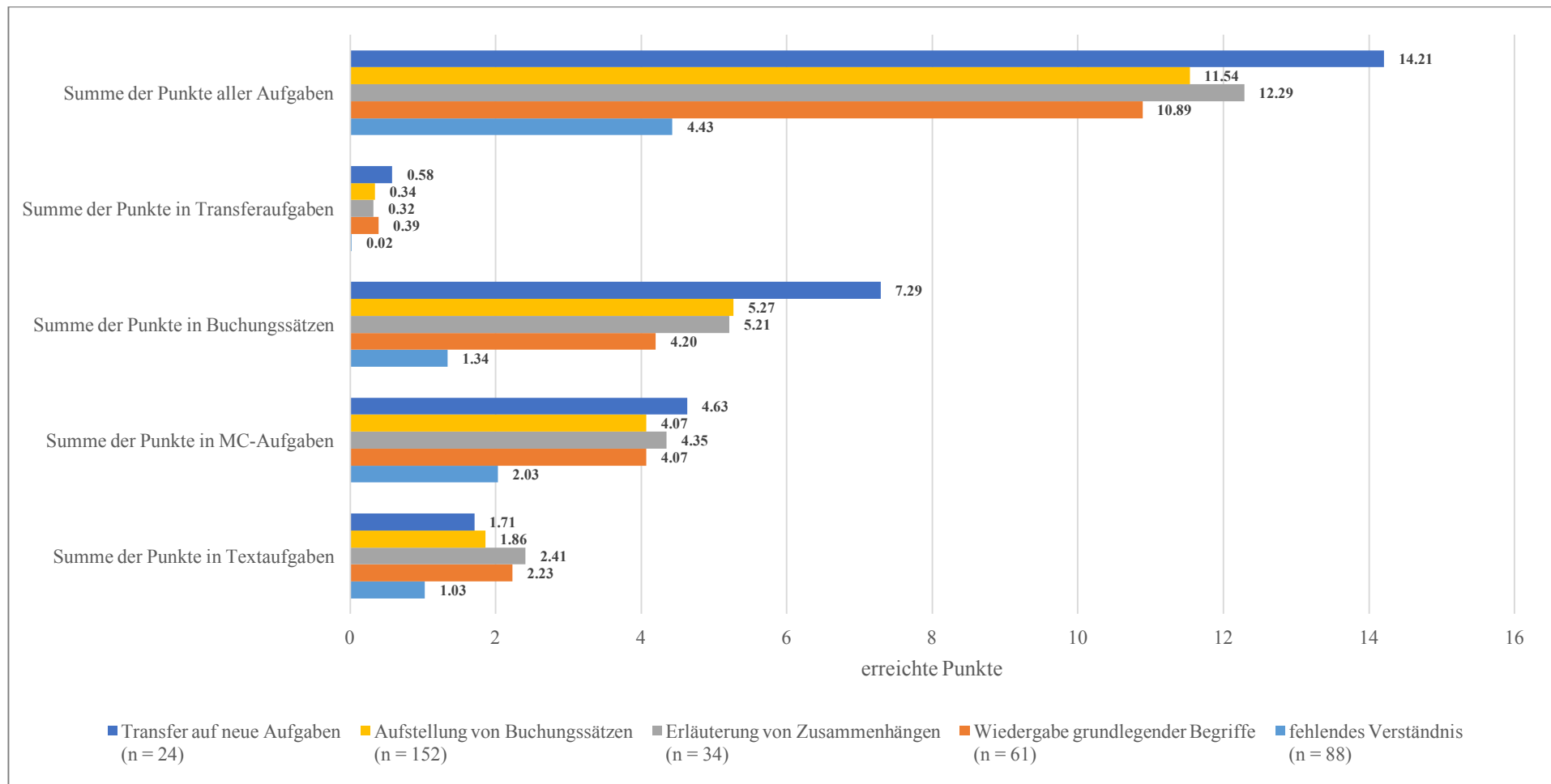
Anhang A5 Mittelwertvergleich der Selbsteinschätzungen nach der Aufgabenbearbeitung für beide Messzeitpunkte ($49 \leq n \leq 56$)

	Messzeitpunkt I		Messzeitpunkt II		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Fehlendes Verständnis	3.44	1.46	2.42	1.20	-4.49	54	.000
Wiedergabe grundlegender Begriffe und Prinzipien	2.89	1.19	3.52	1.13	3.80	55	.000
Erläuterung von Zusammenhängen	2.73	1.15	3.79	1.12	6.85	55	.000
Aufstellung von Buchungssätzen I	3.48	1.53	4.38	.96	4.97	55	.000
Transfer auf neue Aufgaben I	2.55	1.29	3.59	1.16	6.69	55	.000
Aufstellung von Buchungssätzen II	3.20	1.50	4.78	.945	7.71	53	.000
Transfer auf neue Aufgaben II	2.31	1.26	3.43	1.26	7.19	48	.000

Anhang A6 Mittelwertvergleich der Überzeugungen für beide Messzeitpunkte
($27 \leq n \leq 52$)

	Messzeitpunkt I		Messzeitpunkt II		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Textaufgabe I	2.61	1.33	3.41	1.56	3.30	45	.002
Textaufgabe II	1.48	.89	2.68	1.82	3.56	30	.001
Textaufgabe III	3.82	1.11	4.10	1.25	1.41	48	.164
MC-Aufgaben	2.98	1.37	3.96	1.10	5.68	51	.000
Buchungsaufgaben	3.02	1.56	5.07	.97	9.09	43	.000
Transferaufgabe	1.96	1.61	2.74	1.61	2.27	26	.032

Anhang A7 Übersicht der erreichten Punktzahl in Abhängigkeit von der Selbsteinschätzung (SE121)⁴⁷



⁴⁷ Alle Studierenden, die sich auf derselben Niveaustufe einschätzten, wurden zu einer Gruppe zusammengefasst. Anschließend wurde das arithmetische Mittel der Punkte in den jeweiligen Aufgabenbereichen berechnet. Da die Punkteverteilung nicht bei allen Aufgaben normalverteilt war, konnte keine ANOVA zur Signifikanzprüfung eingesetzt werden, stattdessen wurde auf den Kruskal-Wallis-Test als nichtparametrisches Pendant zur Varianzanalyse zurückgegriffen.

Anhang A8 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	382	-.305**	-.487**	-.517**	-.164**	-.561**
Wiedergabe	Korrelation nach Pearson	381	.195**	.307**	.352**	.069	.365**
Erläuterung	Korrelation nach Pearson	382	.214**	.371**	.430**	.171**	.451**
Aufstellung BS	Korrelation nach Pearson	381	.212**	.398**	.496**	.205**	.504**
Transfer	Korrelation nach Pearson	381	.167**	.310**	.461**	.218**	.446**
Anmerkungen		* p < .05; ** p < .01					

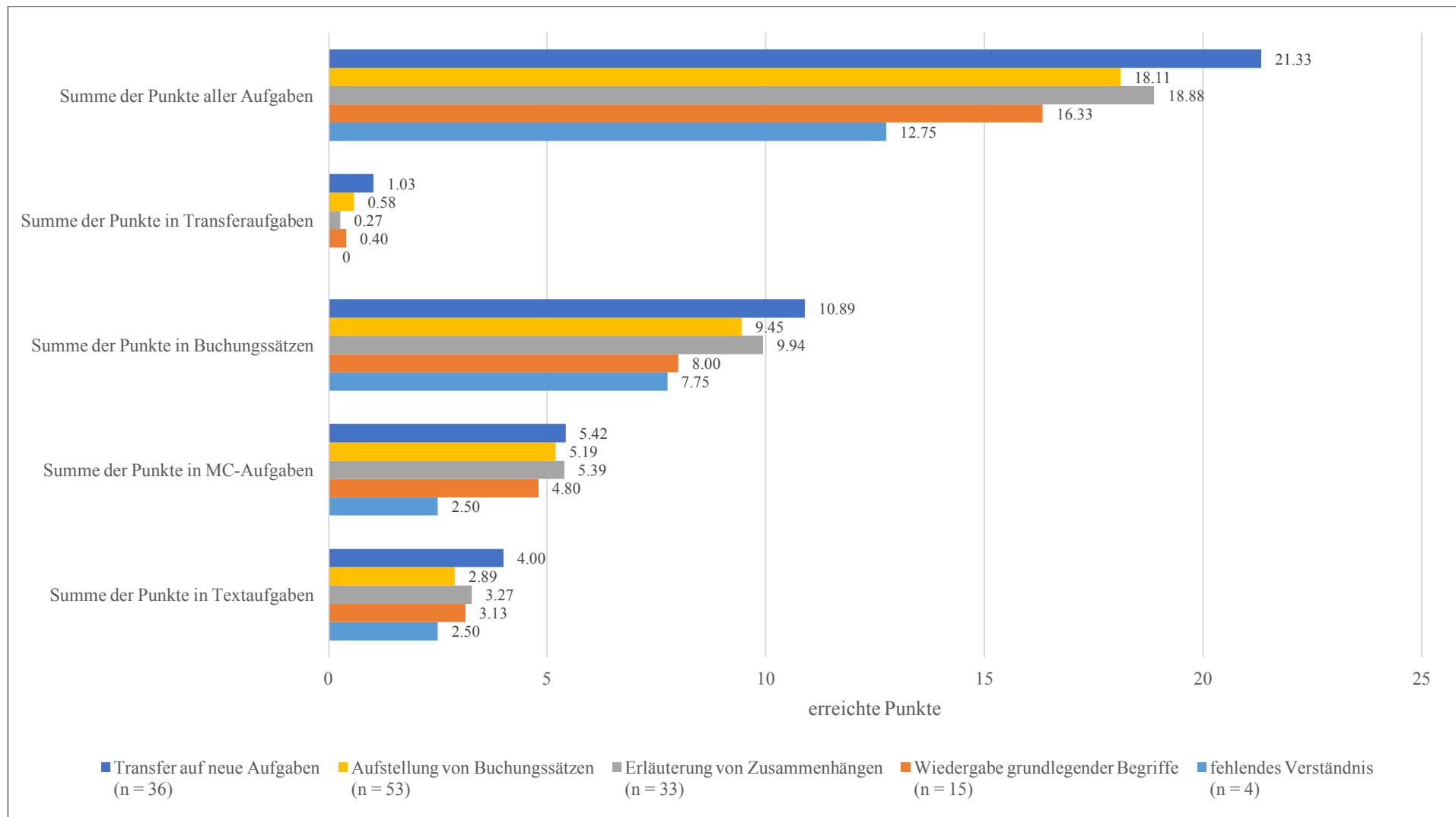
Anhang A9 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	379	-.250**	-.407**	-.412**	-.208**	-.467**
Wiedergabe	Korrelation nach Pearson	377	.269**	.266**	.365**	.090	.381**
Erläuterung	Korrelation nach Pearson	373	.276**	.307**	.406**	.156**	.430**
Aufstellung BS I	Korrelation nach Pearson	376	.282**	.444**	.564**	.265**	.586**
Transfer I	Korrelation nach Pearson	381	.220**	.354**	.490**	.254**	.497**
Aufstellung BS II	Korrelation nach Pearson	368	.269**	.511**	.667**	.267**	.669**
Transfer II	Korrelation nach Pearson	332	.267**	.390**	.507**	.303**	.533**
Anmerkungen		* p < .05; ** p < .01					

Anhang A10 Übersicht über Zusammenhänge zwischen Überzeugungen und der Performanz zum ersten Messzeitpunkt

Überzeugung/Performanz		N	Textaufgabe I	Textaufgabe II	Textaufgabe III	Σ MC- Aufgaben	Σ Buchungs- aufgabe	Σ Transfer- aufgabe	Σ aller Aufgaben
Textaufgabe I	Korrelation nach Pearson	308	.198**	.083	.138*	.214**	.270**	.046	.281**
Textaufgabe II	Korrelation nach Pearson	242	.049	.471**	.164*	.057	.111	-.106	.157*
Textaufgabe III	Korrelation nach Pearson	345	.119*	-.001	.242**	.130*	.187*	-.001	.198**
MC-Aufgaben	Korrelation nach Pearson	353	.214**	.145**	.114*	.554**	.510**	.322**	.582**
Buchungsaufgabe	Korrelation nach Pearson	313	.181**	.144*	.136*	.450**	.601**	.208**	.588**
Transferaufgabe	Korrelation nach Pearson	240	.209**	.191**	.026	.354**	.328**	.408**	.408**
Anmerkungen		* p < .05; ** p < .01							

Anhang A11 Übersicht der erreichten Punktzahl in Abhängigkeit von der Selbsteinschätzung (SE221)



Anhang A12 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum zweiten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	162	-.208**	-.144	-.301**	-.151	-.314**
Wiedergabe	Korrelation nach Pearson	162	.228**	.113	.107	.026	.190*
Erläuterung	Korrelation nach Pearson	162	.191*	.246**	.268**	.111	.313**
Aufstellung BS	Korrelation nach Pearson	162	.103	.191*	.141	.003	.170*
Transfer	Korrelation nach Pearson	162	.218**	.233**	.328**	.131	.354**
Anmerkungen	* p < .05; ** p < .01						

Anhang A13 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum zweiten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	162	-.198*	-.187*	-.260**	-.216**	-.323**
Wiedergabe	Korrelation nach Pearson	163	.297**	.142	.243**	.059	.302**
Erläuterung	Korrelation nach Pearson	163	.270**	.253**	.441**	.141	.437**
Aufstellung BS I	Korrelation nach Pearson	163	.200*	.213**	.438**	.107	.384**
Transfer I	Korrelation nach Pearson	163	.291**	.240**	.414**	.105	.420**
Aufstellung BS II	Korrelation nach Pearson	159	.236**	.324**	.568**	.169*	.501**
Transfer II	Korrelation nach Pearson	153	.244**	.291**	.379**	.307**	.446**
Anmerkungen	* p < .05; ** p < .01						

Anhang A14 Übersicht über Zusammenhänge zwischen Überzeugungen und der Performanz zum zweiten Messzeitpunkt

Überzeugung/Performanz		N	Textaufgabe I	Textaufgabe II	Textaufgabe III	Σ MC- Aufgaben	Σ Buchungs- aufgaben	Σ Trans- feraufgabe	Σ aller Aufgaben
Textaufgabe I	Korrelation nach Pearson	139	.308**	.097	.104	.213*	.220**	-.005	.268*
Textaufgabe II	Korrelation nach Pearson	109	.113	.696**	.054	.153	.232*	.143	.470*
Textaufgabe III	Korrelation nach Pearson	149	.041	.141	.036	.077	.166*	-.014	.156
MC-Aufgaben	Korrelation nach Pearson	156	.139	.119	.243**	.416**	.439**	.221**	.485**
Buchungsaufgabe	Korrelation nach Pearson	141	.170*	.145	.124	.361**	.557**	.216*	.510**
Transferaufgabe	Korrelation nach Pearson	92	.039	.192	-.058	.125	.269**	.404**	.340**
Anmerkungen		* p < .05; ** p < .01							

Anhang A15 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit des Geschlechts

	Weiblich		Männlich		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	(259 ≤ n ≤ 260)		(99 ≤ n ≤ 100)				
Textaufgaben	1.79	1.13	1.75	1.27	.261	357	.795
MC-Aufgaben	1.32	.95	1.18	.83	1.258	358	.213
Buchungsaufgaben	1.61	1.21	1.40	1.07	1.532	357	.126
Transferaufgabe	1.55	1.31	1.77	1.27	-1.418	357	.157
Initiierungszeitpunkt II	(229 ≤ n ≤ 258)		(92 ≤ n ≤ 100)				
Textaufgaben	1.19	1.01	1.23	.91	-.359	356	.720
MC-Aufgaben	1.50	1.13	1.36	.93	1.209	352	.228
Buchungsaufgaben I	1.14	.98	1.18	.92	-.409	355	.683
Transferaufgabe I	1.21	1.23	1.47	1.21	-1.817	354	.070
Buchungsaufgaben II	.82	.79	1.15	.89	-3.153	348	.002
Transferaufgabe II	.93	1.09	1.42	1.26	-3.242	319	.001
Überzeugungen	(156 ≤ n ≤ 238)		(74 ≤ n ≤ 95)				
Textaufgabe I	1.21	1.25	1.47	1.38	-1.545	289	.124
Textaufgabe II	.37	.84	.61	1.04	-1.751	228	.083
Textaufgabe III	1.39	1.01	1.51	1.12	-.971	323	.332
MC-Aufgaben	1.35	1.05	1.22	.93	1.055	331	.292
Buchungsaufgaben	.93	.85	1.27	1.07	-2.682	295	.008
Transferaufgabe	.83	1.19	1.13	1.45	-1.598	231	.113

Anhang A16 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit des Geschlechts

	Weiblich		Männlich		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	n = 73		n = 40				
Textaufgaben	1.72	.98	1.77	.98	-.283	111	.778
MC-Aufgaben	1.19	.87	1.25	.88	-.361	111	.719
Buchungsaufgaben	1.11	1.06	.92	.77	1.002	111	.318
Transferaufgabe	2.26	1.13	2.46	1.33	-.868	111	.387
Initiierungszeitpunkt II	(70 ≤ n ≤ 73)		(38 ≤ n ≤ 40)				
Textaufgaben	1.27	.94	1.32	.90	-.291	111	.772
MC-Aufgaben	1.57	.94	1.51	1.06	.324	111	.747
Buchungsaufgaben I	1.14	.92	.79	.52	2.557	111	.012
Transferaufgabe I	2.20	1.14	2.39	1.18	-.817	111	.416
Buchungsaufgaben II	.80	.72	.79	.51	.049	110	.961
Transferaufgabe II	1.93	1.23	1.80	1.07	.547	106	.586
Überzeugungen	(38 ≤ n ≤ 61)		(29 ≤ n ≤ 39)				
Textaufgabe I	1.62	1.32	1.53	1.43	.327	94	.745
Textaufgabe II	.86	.99	.75	1.05	.466	73	.642
Textaufgabe III	1.23	.91	1.31	1.08	-.394	101	.694
MC-Aufgaben	1.27	.83	.88	.80	2.321	105	.022
Buchungsaufgaben	.81	.75	.58	.54	1.633	97	.106
Transferaufgabe	1.39	1.48	1.41	1.72	-.065	65	.948

Anhang A17 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit der Hochschulzugangsberechtigung

	HZB allgemein		HZB sonst		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	(276 ≤ n ≤ 277)		(n = 87)				
Textaufgaben	1.71	1.18	2.06	1.18	-2.420	361	.016
MC-Aufgaben	1.28	.97	1.37	.90	-.731	362	.465
Buchungsaufgaben	1.58	1.20	1.48	1.13	.689	361	.491
Transferaufgabe	1.48	1.22	2.14	1.48	-3.750	361	.000
Initiierungszeitpunkt II	(243 ≤ n ≤ 274)		(79 ≤ n ≤ 87)				
Textaufgaben	1.15	1.00	1.38	.97	-1.853	358	.065
MC-Aufgaben	1.43	1.13	1.58	1.02	-1.087	354	.278
Buchungsaufgaben I	1.13	1.00	1.27	.90	-1.154	357	.249
Transferaufgabe I	1.23	1.23	1.54	1.27	-2.026	356	.044
Buchungsaufgaben II	.91	.85	.96	.89	-.510	351	.610
Transferaufgabe II	1.02	1.17	1.30	1.14	-1.876	320	.062
Überzeugungen	(182 ≤ n ≤ 257)		(48 ≤ n ≤ 80)				
Textaufgabe I	1.23	1.24	1.62	1.52	-1.875	291	.064
Textaufgabe II	.37	.80	.88	1.35	-2.497	231	.015
Textaufgabe III	1.40	1.04	1.57	1.10	-1.236	326	.217
MC-Aufgaben	1.33	1.04	1.27	1.01	.425	334	.671
Buchungsaufgaben	1.01	.99	1.11	.84	-.734	298	.463
Transferaufgabe	.93	1.36	1.02	1.13	-.449	233	.654

Anhang A18 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit der Hochschulzugangsberechtigung

	HZB allgemein		HZB sonst		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	n = 86		n = 21				
Textaufgaben	1.72	1.02	1.86	.84	-.556	105	.580
MC-Aufgaben	1.22	.85	1.18	.89	.219	105	.827
Buchungsaufgaben	1.01	.98	1.03	.87	-.082	105	.935
Transferaufgabe	2.30	1.23	2.50	1.12	-.679	105	.498
Initiierungszeitpunkt II	(85 ≤ n ≤ 86)		(19 ≤ n ≤ 21)				
Textaufgaben	1.28	.95	1.30	.90	-.086	105	.932
MC-Aufgaben	1.64	.97	1.31	1.00	1.405	105	.163
Buchungsaufgaben I	1.01	.83	.88	.79	.631	105	.529
Transferaufgabe I	2.25	1.16	2.41	1.20	-.533	105	.595
Buchungsaufgaben II	.83	.70	.69	.44	.819	104	.414
Transferaufgabe II	1.85	1.12	2.24	1.32	-1.314	102	.192
Überzeugungen	(57 ≤ n ≤ 83)		(8 ≤ n ≤ 20)				
Textaufgabe I	1.55	1.35	1.90	1.40	-.996	91	.322
Textaufgabe II	.85	1.07	.75	.75	.308	70	.759
Textaufgabe III	1.25	.97	1.23	1.00	.102	97	.919
MC-Aufgaben	1.17	.80	.96	.99	.992	101	.324
Buchungsaufgaben	.75	.74	.73	.43	.129	91	.898
Transferaufgabe	1.23	1.50	2.56	1.76	-2.302	63	.025

Anhang A19 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum ersten Messzeitpunkt in Abhängigkeit der Motivation

	Motivation tief ⁴⁸		Motivation hoch		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	(201 ≤ n ≤ 202)		(n = 165)				
Textaufgaben	1.60	1.13	2.10	1.23	-4.099	364	.000
MC-Aufgaben	1.28	.99	1.29	.89	-.060	365	.952
Buchungsaufgaben	1.47	1.18	1.60	1.17	-1.090	364	.276
Transferaufgabe	1.45	1.21	1.93	1.40	-3.421	364	.001
Initiierungszeitpunkt II	(179 ≤ n ≤ 200)		(150 ≤ n ≤ 164)				
Textaufgaben	1.01	.88	1.46	1.07	-4.290	362	.000
MC-Aufgaben	1.44	1.13	1.46	1.04	-.098	359	.922
Buchungsaufgaben I	1.05	.92	1.24	.98	-1.937	362	.054
Transferaufgabe I	1.00	1.07	1.69	1.36	-5.231	361	.000
Buchungsaufgaben II	.85	.85	.98	.83	-1.390	355	.166
Transferaufgabe II	.75	.97	1.51	1.26	-6.023	327	.000
Überzeugungen	(124 ≤ n ≤ 163)		(108 ≤ n ≤ 159)				
Textaufgabe I	1.04	1.21	1.66	1.34	-4.187	296	.000
Textaufgabe II	.39	.82	.60	1.12	-1.637	232	.103
Textaufgabe III	1.33	1.04	1.58	1.06	-2.151	331	.032
MC-Aufgaben	1.30	.99	1.30	1.03	.059	338	.953
Buchungsaufgaben	1.04	.97	1.04	.92	.030	304	.976
Transferaufgabe	.64	1.09	1.30	1.44	-3.959	234	.000

⁴⁸ Schätzten die Studierenden ihre Motivation auf der sechsstufigen Likert-Skala zwischen 1 und 3 ein, wurden sie der Gruppe mit tiefer ausgeprägter Motivation zugeordnet. Bei Angaben zwischen 4 und 6 erfolgte die Zuteilung zur Gruppe von Studierenden mit höher ausgeprägter Motivation.

Anhang A20 Mittelwertvergleich der Akkuratheit der Selbsteinschätzungen zum zweiten Messzeitpunkt in Abhängigkeit der Motivation

	Motivation tief ⁴⁹		Motivation hoch		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I	n = 67		n = 87				
Textaufgaben	1.71	1.04	1.82	1.04	-.675	152	.500
MC-Aufgaben	1.20	.77	1.26	.96	-4.29	152	.669
Buchungsaufgaben	1.08	.86	.99	1.03	.548	152	.584
Transferaufgabe	2.16	1.10	2.64	1.23	-2.531	152	.012
Initiierungszeitpunkt II	(62 ≤ n ≤ 67)		(86 ≤ n ≤ 87)				
Textaufgaben	1.25	.85	1.40	1.04	-.954	152	.341
MC-Aufgaben	1.42	.91	1.53	1.00	-.659	152	.511
Buchungsaufgaben I	1.01	.79	1.01	.86	.014	152	.989
Transferaufgabe I	2.04	1.10	2.61	1.21	-3.013	152	.003
Buchungsaufgaben II	.94	.68	.76	.66	1.714	149	.089
Transferaufgabe II	1.82	1.16	2.15	1.22	-1.659	147	.099
Überzeugungen	(37 ≤ n ≤ 63)		(54 ≤ n ≤ 85)				
Textaufgabe I	1.61	1.33	1.78	1.37	-.689	129	.492
Textaufgabe II	.77	1.00	.98	1.14	-1.008	103	.316
Textaufgabe III	1.16	.94	1.39	1.00	-1.385	140	.168
MC-Aufgaben	1.08	.81	1.23	.91	-1.056	146	.293
Buchungsaufgaben	.84	.684	.68	.70	1.391	132	.167
Transferaufgabe	1.46	1.41	1.69	1.70	-.679	89	.499

⁴⁹ Die Gruppenbildung erfolgte analog zur in Anhang A19 beschriebenen Vorgehensweise.

Anhang A21 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen vor der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	56	-.164	-.442**	-.544**	-.273*	-.532**
Wiedergabe	Korrelation nach Pearson	56	.240	.296*	.287*	.174	.345**
Erläuterung	Korrelation nach Pearson	56	.195	.405**	.446**	.289*	.476**
Aufstellung BS	Korrelation nach Pearson	56	.096	.385**	.371**	.325**	.408**
Transfer	Korrelation nach Pearson	56	.035	.504**	.529**	.394**	.529**
Anmerkungen	* $p < .05$; ** $p < .01$						

Anhang A22 Übersicht über Zusammenhänge zwischen Selbsteinschätzungen nach der Aufgabenbearbeitung und der Performanz zum ersten Messzeitpunkt

Selbsteinschätzung/Performanz		N	Σ Textaufgaben	Σ MC-Aufgaben	Σ Buchungsaufgaben	Σ Transferaufgabe	Σ aller Aufgaben
Fehlendes Verständnis	Korrelation nach Pearson	56	-.263	-.577**	-.576**	-.489**	-.651**
Wiedergabe	Korrelation nach Pearson	56	.307*	.340*	.342**	.351**	.436**
Erläuterung	Korrelation nach Pearson	56	.176	.452**	.447**	.313*	.488**
Aufstellung BS I	Korrelation nach Pearson	56	.245	.634**	.594**	.447**	.666**
Transfer I	Korrelation nach Pearson	56	.081	.646**	.620**	.488**	.648**
Aufstellung BS II	Korrelation nach Pearson	55	.133	.637**	.667**	.470**	.689**
Transfer II	Korrelation nach Pearson	49	.158	.527**	.617**	.474**	.631**
Anmerkungen	* $p < .05$; ** $p < .01$						

Anhang A23 Mittelwertvergleich der Performanz in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen insgesamt (N = 56)

	Akkuratere SE		Inakkuratere SE		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Initiierungszeitpunkt I							
Textaufgaben	3.21	2.35	4.04	2.33	-1.326	54	.190
MC-Aufgaben	4.76	1.79	5.89	1.09	-2.883	54	.006
Buchungsaufgaben	9.55	2.71	10.67	1.47	-1.896	54	.006
Transferaufgabe	1.00	1.65	1.07	1.77	-.162	54	.872
Σ alle Aufgaben	18.52	6.49	21.67	4.42	-2.108	54	.040
Initiierungszeitpunkt II							
Textaufgaben	2.75	1.88	4.34	2.54	-2.702	55	.009
MC-Aufgaben	4.75	1.94	5.83	.85	-2.705	55	.010
Buchungsaufgaben	9.36	2.79	10.79	1.18	-2.515	55	.017
Transferaufgabe	1.00	1.68	1.03	1.72	-.077	55	.939
Σ alle Aufgaben	17.86	6.27	22.00	4.37	-2.903	55	.005
Überzeugungen							
Textaufgabe I	.31	.62	.54	.79	-1.184	52	.242
Textaufgabe II	1.36	2.08	1.21	1.81	.219	40	.828
Textaufgabe III	1.92	.83	1.90	.67	.097	51	.923
MC-Aufgaben	5.36	1.62	5.19	1.57	.400	53	.691
Buchungsaufgaben	9.83	2.86	10.44	1.34	-.953	48	.348
Transferaufgabe	.90	1.65	1.47	1.98	-.979	37	.334
Σ alle Aufgaben	18.14	5.94	21.72	5.02	-2.462	55	.017

Anhang A24 Deskriptive Ergebnisse zu den Items der Lernstrategien und Lernzeit zu beiden Messzeitpunkten

		MZ 1 (354 ≤ n ≤ 372)		MZ 2 (148 ≤ n ≤ 160)	
Lernstrategie		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Veranstaltungsbesuch	LS1: Besuch Vorlesung	2.38	.71	2.15	.75
	LS2: Besuch Übung	2.67	.57	2.11	.71
	LS3: Besuch Tutorien	2.50	.68	2.33	.68
Vor- und Nachbereitung	LS4: Vorbereitung Vorlesung	1.86	1.32	2.32	1.53
	LS5: Vorbereitung Übung	2.38	1.55	2.74	1.65
	LS6: Vorbereitung Tutorien	2.93	1.77	3.21	1.78
	LS7: Nachbereitung Vorlesung	3.26	1.70	4.20	1.75
	LS8: Nachbereitung Übung	2.99	1.70	4.03	1.80
	LS9: Nachbereitung Tutorien	3.17	1.78	4.29	1.71
Lösen von Übungsaufgaben	LS10: eigenständige Lösung der Übungsaufgaben	3.06	1.74	4.03	1.62
	LS11: Lösung der Übungsaufgaben mit Kommilitonen	1.82	1.24	2.23	1.54
	LS12: Nutzung der Vorlesungsunterlagen	3.06	1.70	3.89	1.69
	LS13: Zuwendung schwierigere Aufgaben	2.74	1.52	3.66	1.37
	LS14: Wiederholung Übungsaufgaben	2.43	1.47	3.39	1.46
Umgang mit Schwierigkeiten	LS15: Rückstellung von Übungsaufgaben	3.56	1.70	3.73	1.54
	LS16: Rat von Kommilitonen	3.30	1.71	3.84	1.70
	LS17: Nutzung der vorlesungsbegleitenden Literatur	3.36	1.69	4.17	1.75
	LS18: Bearbeitung einfacher Aufgaben bei Schwierigkeiten	3.68	1.60	3.99	1.38
Zeitaufwand		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Zeitaufwand	ZA1: relativer Zeitaufwand	2.39	1.22	3.30	1.24
	ZA2: Aufwand für Vor- und Nachbereitung der Vorlesung [Minuten]	53.69	63.45	99.70	132.62
	ZA3: Aufwand für Vor- und Nachbereitung der Übung [Minuten]	41.47	68.96	69.99	66.08
	ZA4: Aufwand für Vor- und Nachbereitung der Tutorien [Minuten]	47.42	64.17	83.67	76.19

Anhang A25 Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) am ersten Messzeitpunkt

Akkuratheit nach Aufgabe	Kategorie	akkurate SE		inakkurate SE		<i>t</i>	<i>df</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Textaufgabe	Veranstaltungsbesuch	2.43	.55	2.55	.46	1.908	363	.058
	Vor- und Nachbereitung	2.40	1.22	2.90	1.23	3.567	366	.000
	Bearbeitung der Aufgaben	2.28	1.04	2.76	1.08	3.877	364	.000
	Umgang mit Schwierigkeiten	3.32	1.36	3.51	1.22	1.283	369	.201
MC-Aufgaben	Veranstaltungsbesuch	2.48	.48	2.54	.49	1.205	364	.229
	Vor- und Nachbereitung	2.72	1.25	2.78	1.25	.451	367	.653
	Bearbeitung der Aufgaben	2.62	1.06	2.62	1.11	-.059	365	.953
	Umgang mit Schwierigkeiten	3.40	1.22	3.50	1.29	.696	370	.487
Buchungsaufgaben	Veranstaltungsbesuch	2.52	.49	2.52	.49	-.029	363	.977
	Vor- und Nachbereitung	2.55	1.29	2.92	1.19	2.817	366	.005
	Bearbeitung der Aufgaben	2.48	1.10	2.73	1.07	2.230	364	.026
	Umgang mit Schwierigkeiten	3.37	1.32	3.53	1.21	1.218	369	.224
Transferaufgabe	Veranstaltungsbesuch	2.48	.50	2.55	.47	1.233	363	.218
	Vor- und Nachbereitung	2.50	1.19	3.02	1.25	4.052	366	.000
	Bearbeitung der Aufgaben	2.37	1.00	2.87	1.12	4.473	364	.000
	Umgang mit Schwierigkeiten	3.35	1.27	3.56	1.25	1.632	369	.104

Anhang A26 Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzungen nach der Aufgabenbearbeitung (Initiierungszeitpunkt II) am ersten Messzeitpunkt

Akkuratheit nach Aufgabe	Kategorie	akkurate SE		inakkurate SE		<i>t</i>	<i>df</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Textaufgabe	Veranstaltungsbesuch	2.53	.466	2.50	.50	-.563	360	.573
	Vor- und Nachbereitung	2.50	1.18	3.00	1.27	3.896	363	.000
	Bearbeitung der Aufgaben	2.45	1.09	2.79	1.18	3.059	361	.002
	Umgang mit Schwierigkeiten	3.35	1.31	3.57	1.20	1.615	366	.107
MC-Aufgaben	Veranstaltungsbesuch	2.49	.50	2.53	.48	.922	357	.357
	Vor- und Nachbereitung	2.69	1.27	2.80	1.23	.871	360	.384
	Bearbeitung der Aufgaben	2.56	1.08	2.67	1.10	.952	358	.342
	Umgang mit Schwierigkeiten	3.46	1.23	3.46	1.28	-.021	363	.984
Buchungsaufgaben I	Veranstaltungsbesuch	2.54	.50	2.50	.46	-.817	360	.415
	Vor- und Nachbereitung	2.56	1.21	2.97	1.26	3.130	363	.002
	Bearbeitung der Aufgaben	2.55	1.07	2.71	1.11	1.394	361	.164
	Umgang mit Schwierigkeiten	3.38	1.28	3.55	1.23	1.355	366	.176
Transferaufgabe I	Veranstaltungsbesuch	2.49	.49	2.55	.48	1.110	359	.268
	Vor- und Nachbereitung	2.52	1.21	3.12	1.22	4.589	362	.000
	Bearbeitung der Aufgaben	2.38	1.06	2.99	1.04	5.317	360	.000
	Umgang mit Schwierigkeiten	3.33	1.27	3.66	1.22	2.485	365	.013
Buchungsaufgaben II	Veranstaltungsbesuch	2.52	.51	2.53	.42	.156	357	.876
	Vor- und Nachbereitung	2.65	1.23	2.99	1.23	2.487	357	.013
	Bearbeitung der Aufgaben	2.50	1.07	2.85	1.09	2.954	355	.003
	Umgang mit Schwierigkeiten	3.33	1.24	3.62	1.26	2.137	359	.330
Transferaufgabe II	Veranstaltungsbesuch	2.52	.46	2.56	.47	.786	330	.432
	Vor- und Nachbereitung	2.56	1.16	3.20	1.19	4.581	329	.000
	Bearbeitung der Aufgaben	2.40	1.03	3.12	1.03	5.948	328	.000
	Umgang mit Schwierigkeiten	3.30	1.27	3.78	1.11	3.513	330	.001

Anhang A27 Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzung vor der Aufgabenbearbeitung (Initiierungszeitpunkt I) am zweiten Messzeitpunkt

Akkuratheit nach Aufgabe	Kategorie	akkurate SE		inakkurate SE		<i>t</i>	<i>df</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Textaufgabe	Veranstaltungsbesuch	2.25	.58	2.18	.55	-.683	157	.495
	Vor- und Nachbereitung	3.41	1.41	3.46	1.12	.240	157	.811
	Bearbeitung der Aufgaben	3.16	1.17	3.52	.88	1.775	155	.082
	Umgang mit Schwierigkeiten	4.00	1.15	3.87	1.00	-.687	155	.493
MC-Aufgaben	Veranstaltungsbesuch	2.18	.58	2.20	.54	.210	157	.834
	Vor- und Nachbereitung	3.52	1.22	3.40	1.18	-.606	157	.546
	Bearbeitung der Aufgaben	3.39	1.03	3.46	.93	.399	155	.690
	Umgang mit Schwierigkeiten	3.89	1.18	3.91	.93	.068	155	.946
Buchungsaufgaben	Veranstaltungsbesuch	2.21	.50	2.18	.61	-.423	157	.673
	Vor- und Nachbereitung	3.59	1.14	3.30	1.24	-1.545	157	.124
	Bearbeitung der Aufgaben	3.42	.91	3.44	1.04	.134	155	.893
	Umgang mit Schwierigkeiten	3.80	1.12	4.01	.92	.1248	155	.214
Transferaufgabe	Veranstaltungsbesuch	2.05	.60	2.23	.54	1.585	157	.115
	Vor- und Nachbereitung	2.79	1.18	3.60	1.15	3.499	157	.001
	Bearbeitung der Aufgaben	3.18	.95	3.48	.97	1.488	155	.139
	Umgang mit Schwierigkeiten	3.61	1.19	3.97	.99	1.670	155	.097

Anhang A28 Mittelwertvergleich der Lernstrategien in Abhängigkeit von der Akkuratheit der Selbsteinschätzung nach der Aufgabenbearbeitung (Initiierungszeitpunkt II) am zweiten Messzeitpunkt

Akkuratheit nach Aufgabe	Kategorie	akkurate SE		inakkurate SE		<i>t</i>	<i>df</i>	<i>p</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Textaufgabe	Veranstaltungsbesuch	2.20	.61	2.20	.52	-.058	158	.954
	Vor- und Nachbereitung	3.33	1.18	3.54	1.20	1.132	158	.259
	Bearbeitung der Aufgaben	3.26	1.11	3.54	.85	1.695	156	.093
	Umgang mit Schwierigkeiten	4.07	1.08	3.82	1.01	-1.496	156	.137
MC-Aufgaben	Veranstaltungsbesuch	2.11	.58	2.24	.54	1.357	158	.177
	Vor- und Nachbereitung	3.55	1.12	3.42	1.23	-.666	158	.507
	Bearbeitung der Aufgaben	3.54	.95	3.39	.98	-.908	156	.365
	Umgang mit Schwierigkeiten	3.83	1.22	3.95	.96	.668	156	.505
Buchungsaufgaben I	Veranstaltungsbesuch	2.14	.55	2.26	.56	1.381	158	.169
	Vor- und Nachbereitung	3.42	1.18	3.49	1.21	.365	158	.716
	Bearbeitung der Aufgaben	3.47	.91	3.40	1.04	-.429	156	.669
	Umgang mit Schwierigkeiten	3.90	1.07	3.94	1.06	.240	156	.811
Transferaufgabe I	Veranstaltungsbesuch	2.18	.60	2.20	.54	.269	158	.788
	Vor- und Nachbereitung	3.25	1.25	3.52	1.18	1.149	158	.252
	Bearbeitung der Aufgaben	3.20	1.03	3.50	.95	1.614	156	.108
	Umgang mit Schwierigkeiten	4.01	1.17	3.89	1.01	-.630	156	.529
Buchungsaufgaben II	Veranstaltungsbesuch	2.17	.57	2.24	.54	.764	156	.446
	Vor- und Nachbereitung	3.49	1.18	3.45	1.21	-.190	155	.849
	Bearbeitung der Aufgaben	3.47	.93	3.35	1.02	-.722	153	.471
	Umgang mit Schwierigkeiten	3.82	1.01	4.04	1.08	1.358	153	.176
Transferaufgabe II	Veranstaltungsbesuch	2.25	.59	2.21	.54	-.340	150	.734
	Vor- und Nachbereitung	3.18	1.14	3.59	1.19	1.942	150	.054
	Bearbeitung der Aufgaben	3.24	1.00	3.47	.94	1.347	150	.180
	Umgang mit Schwierigkeiten	3.83	1.03	3.93	1.06	.559	150	.577